

Gane Samb LO

Convergence Vague (IA)

-

Suites de Vecteurs Aléatoires

*Société Africaine de Probabilités et de
Statistiques (SPAS) Editions Series.
Saint-Louis, SENEGAL - Calgary,
Alberta. 2016.*

DOI : <http://dx.doi.org/10.16929/sbs/2016.0002>

ISBN 978-2-9559183-2-6

SPAS Series Books

Advisers

List of published books

Library of Congress Cataloging-in-Publication Data

Gane Samb LO, 1958-

Convergence Vague (IA). Suites de Vecteurs Aléatoires.

SPAS Editions, 2016.

Copyright ©Socit Africaine de Probabilit et de Statistiques (SPAS).

DOI : 10.16929/sbs/2016.0002

ISBN 978-2-9559183-2-6

Auteur : Gane Samb LO

Emails:

gane-samb.lo@ugb.edu.sn, ganesamblo@ganesamblo.net.

Url's:

www.ganesamblo@ganesamblo.net

www.statpas.net/cva.php?email.ganesamblo@yahoo.com.

Affiliations.

Affiliation principale : Universit Gaston Berger, UGB, SENEGAL.

African University of Sciences and Technology, AUST, ABuja, Nigeria.

Chercheur associé au : LSTA, Universit Pierre et Marie, Paris VI, France.

L'auteur enseigne ou a enseigné au niveau Master dans les universités suivantes:

Saint-Louis, Sngal (UGB)

Banjul, Gambi (TUG)

Bamako, Mali (USTTB)

Ouagadougou - Burkina Faso (UJK)

African Institute of Mathematical Sciences, Mbour, SENEGAL, AIMS.

Franceville, Gabon

Ddicaces.

A mon pouse Mbaye Ndaw Fall, ma compagne depuis des dcades

Manifestation de reconnaissance de Soutien Financier et Matériel.

L'auteur manifeste sa gratitude au Centre d'Excellence de la Banque Mondiale pour les Mathématiques, l'Informatique et les TIC (CEA-MITIC) pour un support continu de ses projets dans les années 2014, 2015 and 2016. Il remercie aussi les autorités de l'Universit Gaston Berger pour le soutien permanent sous toutes ses formes.

**Convergence Vague (IA). Suites de Vecteurs
Aléatoires**

ABSTRACT. (English) This monograph aims at presenting the core weak convergence theory for sequences of random vectors with values in \mathbb{R}^k . In some places, a more general formulation in metric spaces is provided. It lays out the necessary foundation that paves the way to applications in particular subfields of the theory. In particular, the needs of Asymptotic Statistics are addressed. A whole chapter is devoted to weak convergence in \mathbb{R} where specific tools, for example for handling weak convergence of sequences using independent and identically distributed random variables such that the Renyi's representations by means of standard uniform or exponential random variables, are stated. The function empirical process is presented as a powerful tool for solving a considerable number of asymptotic problems in Statistics. The text is written in a self-contained approach whith the proofs of all used results at the exception of the general Skorohod-Wichura Theorem.

(Français) Cet ouvrage a l'ambition de présenter le noyau dur de la théorie de la convergence vague de suite de vecteurs aléatoires dans \mathbb{R}^k . Autant que possible, dans certaines situations, la théorie générale dans des espaces métriques est donnée. Il prépare la voie à une spécialisation dans certains sous-domaines de la convergence vague. En particulier, les besoins de la statistique asymptotique ont été satisfaits. Un chapitre de l'ouvrage concerne la convergence vague dans \mathbb{R} avec des outils spécifiques, par exemple, pour étudier les suites de variables aléatoires indépendantes et identiquement distribuées tels que la représentation de Renyi au moyen de variables aléatoires uniformes ou exponentielles standard. Le processus empirique fonction est introduit comme un outil puissant pour étudier des problèmes asymptotiques en Statistiques. Le texte est rédigé dans une approche auto-citante avec toutes les preuves des résultats utilisés, à l'exception du Théorème de Skorohod-Wichura.

Keywords. Convergence vague; convergence en distribution; Thorme Portmanteau; Caractrisation d'une loi de probabilit; fonction de distribution, fonction de rpartition; fonction caratristiques; densit de probabilit; Marches alatoires; Processus empirique; Loi Multinomiale; Compacit Relative; Tension Asymptotique; Tension uniform; Thorme de la transformation Continue; Reprsentation de Renyi et de Malmquist; Statistiques d'ordre; Mthodes Delta Multivariaries; Proces-sus empirique fonctionnel.

AMS 2010 Classification Subjects : 60XXX; 62G30

Contents

Chapter 1. Revue de convergence vague dans \mathbb{R}^k	3
1. Introduction	3
2. Convergence vague sur \mathbb{R}^k	3
3. Exemples de convergence vague dans \mathbb{R}	6
4. Exemples de convergence dans \mathbb{R}^k	17
5. Principe d'invariance	27
Chapter 2. Théorie de la Convergence Vague	31
1. Introduction	31
2. Définition, Unicité et Théorème Portmanteau	31
3. Transformations continues	40
4. Cas particulier de \mathbb{R}^k , $k \geq 1$	41
5. Théorème de Scheffé	51
6. Convergence vague et convergence en probabilité sur le même espace de probabilité	54
7. Annexe	60
Chapter 3. Tension uniforme et tension asymptotique	71
1. Introduction	71
2. Tension	78
3. Théorème de compacité de Prohorov dans \mathbb{R}^k .	82
4. Applications	88
Chapter 4. Outils Particuliers pour la Convergence Vague dans \mathbb{R}	91
1. Inverses généralisées des fonctions monotones	91
2. Représentation uniform et exponentielles de Renyi	104
Chapter 5. Le processus empirique fonctionnel comme outil général en statistique asymptotique	113
1. Utilisation du petit o et du grand O	113
2. Méthodes Delta	130
3. Utilisation du Processus Empirique Fonctionnel en Statistique Asymptotique	135
Chapter 6. Théorie des fonctions et éléments d'analyse réelle à travers des exercices	147

1. Revue des limites dans $\overline{\mathbb{R}}$. Ce que nous ne devons pas ignorer des limites.	147
2. Miscellaneuous facts	159
Bibliography	161

Ce texte fait partie d'une série dont l'ambition est de brasser une grande partie des probabilités à des fins pédagogiques. Ces textes permettront aux apprenant de se former tous seuls.

Ils pourront constituer pour les professeurs de documents de cours et d'exercices. Pour les plus ambitieux, ils seront une base de départ pour des textes plus avancés et personnalisés. Ils sont mis gracieusement à la disposition des apprenants et des maîtres.

Nos ouvrages sont rangés dans trois catégories :

Une catégorie d'initiation pour les débutants. Il s'agit d'ouvrages souvent accessibles dès la première année d'université et ne demandant pas de pré-requis en mathématiques avancées. Les ouvrages de probabilités élémentaires et de Statistiques élémentaires sont à ranger dans cette catégorie. Cette initiation précède les versions mathématiques de ces théories et préparent les applications de ces dernières.

Une catégorie d'ouvrages d'applications et d'outils. Des étudiants ou des chercheurs dans des disciplines annexes comme l'économie, la médecine, l'hydrologie, la finance, etc. peuvent avoir besoin d'outils assez avancés de la théorie des probabilités ou des statistiques. Ils sont plus intéressés par l'application des outils que leur fondements ou leur développement. Des ouvrages adaptés à ce besoin peuvent être composés. Un parfait exemple est un ouvrage de statistiques mathématiques destinés à des économistes n'ayant pas nécessairement fait la théorie de la mesure.

Une catégorie d'ouvrages spécialisés. Il s'agit d'ouvrages de niveau international rigoureusement écrits avec tous les arguments nécessaires. Ces ouvrages sont conçus pour être consultés partout sur le globe dans leurs versions françaises et anglaises. Ils sont basés sur les ouvrages de base : Théorie de la mesure, Fondements mathématiques des probabilités. À partir de cette base, les ouvrages seront auto-cités, ce qui veut dire que toutes les mathématiques utilisées dans un ouvrage de cette série, en dehors des notions de premier cycle, seront démontrées quelque part dans un ouvrage de la catégorie.

La lecture de ces ouvrages ne requiert alors que le niveau de Licence. Un lecteur capable de lire l'ouvrage de mesure et d'intégration aura les moyens de se spécialiser grâce à cette catégorie, dans beaucoup de branches des probabilités et des statistiques.

Sans insister, nous dirons que nous n'incluons la **catégorie des ouvrages de recherche** qui sont et seront disponibles. Ces ouvrages partagent la même orientation, à savoir qu'ils sont écrits pour être lus après les cours fondamentaux de mesure et des probabilités mathématiques.

Nos collaborateurs et anciens élèves sont priés de faire vivre la chaîne de sorte que le centre de Saint-Louis, donc le Sénégal et l'Afrique, soit un véritable creuset et une grande école de mathématiques, de probabilités et de statistique.

CHAPTER 1

Revue de convergence vague dans \mathbb{R}^k

1. Introduction

Dans ce chapitre, nous allons voir que les lecteurs, pour la plupart d'entre eux, ont déjà étudié plusieurs exemples de convergence vague. Ce qui leur manque peut être, c'est la cohérence mathématique de cette théorie et sa place dans le cadre général des convergences. Ces manquements seront comblés dans ce cours.

Nous allons exhumers des résultats connus de convergence vague rencontrés au cours du parcours déjà effectué. Pour commencer, nous allons émettre une assertion que nous ne serons en mesure de démontrer que dans le chapitre 2, en particulier dans le théorème 3 de ce chapitre.

2. Convergence vague sur \mathbb{R}^k

Commençons par rappeler que d'après l'ouvrage **Base mathématiques des probabilités** [8] que la loi de probabilité d'un vecteur aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{R}^k$ est caractérisée par

(a) sa fonction de répartition:

$$\mathbb{R}^k \ni x \mapsto F_X(x) = \mathbb{P}(X \leq x),$$

(b) sa fonction caractéristique (ici, i est le nombre imaginaire pur vérifiant $i^2 = -1$, et $\langle \cdot, \cdot \rangle$ désigne le produit scalaire classique sur \mathbb{R}^k)

$$\mathbb{R}^k \ni u \mapsto \Phi(u) = E(\exp(i \langle u, X \rangle)),$$

(c) La fonction des moments (si elle existe dans un voisinage du vecteur nul)

$$\mathbb{R}^k \ni u \mapsto \Psi_X(x) = E(\exp(\langle u, X \rangle)).$$

et

(d) par sa dérivée de Radon-Nikodym, (si elle existe), par rapport à une mesure σ -finie ν sur \mathbb{R}^k ,

$$d\mathbb{P}/d\nu = f_X.$$

Il est intéressant que ces caractérisations s'étendent à la convergence vague ainsi, dans un théorème que nous démontrerons plus tard dans le théorème 3 du chapitre 2.

THÉORÈME 1. (*THEOREME - DEFINITION - LEMME*)

Soit une suite vecteurs aléatoires $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ et un autre vecteur aléatoire $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$. Alors les propriétés (a) et (b) suivantes sont équivalentes

(a) Pour tout $u \in \mathbb{R}^k$,

$$\Phi_{X_n}(u) \rightarrow \Phi_X(u) \text{ quand } n \rightarrow +\infty$$

(b) Pour tout point de continuité $x \in \mathbb{R}^k$,

$$F_{X_n}(x) \rightarrow F_X(x) \text{ quand } n \rightarrow +\infty.$$

Si l'un de ces deux points a lieu, alors nous disons que X_n converge vaguement vers X , ou X_n converge en distribution vers X ou X_n converge en loi X , notée

$X_n \rightsquigarrow X$ ou $X_n \xrightarrow{d} X$ ou $X_n \xrightarrow{\mathcal{L}} X$ ou $X_n \xrightarrow{w} X$ ou $X_n \xrightarrow{w} X$ (ici w fait référence au term anglais weak convergence).

Nous avons des conditions suffisantes de convergence vague :

(c) Si de plus, les fonction des moment Ψ_{X_n} sont définies sur B_n , $n \geq 1$ et Ψ_X est définie sur B , où les B_n et B sont des voisinages de 0, et si pour tout $x \in B$,

$$\Psi_{X_n}(x) \rightarrow \Psi_X(x) \text{ quand } n \rightarrow +\infty$$

alors X_n converge vaguement X .

(d) Enfin, si les distributions admettent des dérivées de Radon-Nikodym, c'est-à-dire des densités de probabilités par rapport à la même mesure

σ -finie ν , notées $d\mathbb{P}_n/d\nu = f_{X_n}$ et $d\mathbb{P}/d\nu = f_X$ existent et si pour tout $x \in D_X = \{x, f_X(x) > 0\}$,

$$f_{X_n}(x) \rightarrow f_X(x) \text{ as } n \rightarrow +\infty,$$

alors $X_n \rightsquigarrow X$.

Nous avons ce dernier point.

(e) Supposons que $\{X_n, n \geq 1\} \subset \mathbb{R}^k$ converge vaguement vers $X \in \mathbb{R}^k$, quand $n \rightarrow +\infty$ et soit A une matrice réelle de m lignes et k colonnes avec $m \geq 1$. Alors $\{AX_n, n \geq 1\} \subset \mathbb{R}^m$ converge vaguement vers $AX \in \mathbb{R}^m$.

Remarque. Le point (e) ci-dessus est une conséquence du théorème 7 de la transformation continue établie et prouvée dans le chapitre 2.

En résumé, la convergence en loi sur \mathbb{R}^k a lieu lorsque les fonctions de répartition, les fonctions caractéristiques, les densités de probabilités ou les fonctions des moments convergent vers celle d'une loi de probabilité, pourvu que dans les deux derniers cas, les fonctions en question existent.

Tout cela est énorme. Cela nous sonne l'occasion d'aller directement sur les exemples connus de convergence de ces fonctions et d'en ajouter de nouveaux.

Avant d'aller plus loin, nous aurons souvent besoin de cet outil intéressant pour obtenir la convergence vague à partir de la convergence de fonction caractéristique.

PROPOSITION 1. Crière de Wold. La suite de vecteurs aléatoires $\{X_n, n \geq 1\} \subset \mathbb{R}^k$ converge vaguement vers $X \in \mathbb{R}^k$, quand $n \rightarrow +\infty$ si et seulement si pour $a \in \mathbb{R}^k$, la suite $\{<a, X_n>, n \geq 1\} \subset \mathbb{R}$ converge vaguement vers $X \in \mathbb{R}$ quand $n \rightarrow +\infty$.

Preuve. La preuve est rapide. Elle utilise les notations antérieures. Supposons que X_n converge faiblement vers X dans \mathbb{R}^k quand $n \rightarrow +\infty$. En utilisant la convergence des fonctions caractéristiques, nous avons pour tout $u \in \mathbb{R}^k$,

$$\mathbb{E}(\exp(i < X_n, u >)) \rightarrow \mathbb{E}(\exp(i < X, u >)) \text{ quand } n \rightarrow +\infty.$$

Il s'en suit que $a \in \mathbb{R}^k$ et pour tout $t \in \mathbb{R}$, nous avons

$$(2.1) \quad \mathbb{E}(\exp(it \langle X_n, a \rangle)) \rightarrow \mathbb{E}(\exp(it \langle X, a \rangle)) \text{ quand } n \rightarrow +\infty.$$

c-a-d que, en prenant $u = ta$ dans la formule précédente, et en notant $Z_n = \langle X_n, a \rangle$ et $Z = \langle X, a \rangle$,

$$\mathbb{E}(\exp(itZ_n)) \rightarrow \mathbb{E}(\exp(itZ)) \text{ quand } n \rightarrow +\infty.$$

Cela signifie que $Z_n \rightsquigarrow Z$, qui est égal $\langle a, X \rangle$, converge vaguement vers $\langle a, X \rangle$.

Inversement, supposons que pour tout $a \in \mathbb{R}^k$, la suite $\{\langle a, X_n \rangle, n \geq 1\} \subset \mathbb{R}$ converge vaguement vers $X \in \mathbb{R}$ as $n \rightarrow +\infty$. Alors, en prenant $t = 1$ dans (4.14), nous obtenons pour tout $a = u \in \mathbb{R}^k$,

$$\mathbb{E}(\exp(i \langle X, u \rangle)) \rightarrow \mathbb{E}(\exp(i \langle X, u \rangle)) \text{ quand } n \rightarrow +\infty.$$

ce qui signifie que $X_n \rightsquigarrow +\infty$ quand $n \rightarrow +\infty$.

3. Exemples de convergence vague dans \mathbb{R}

3.1. Convergence de la loi hypergéométrique vers la loi binomiale. Soit X_N suivant une loi hypergéométrique $\mathcal{H}(N, M, n)$ avec $M/N \rightarrow p$, $N \rightarrow \infty$, n restant fixe. Alors X_N tend en loi vers une variable aléatoire X suivant une loi binomiale $\mathcal{B}(n, p)$.

PREUVE. Pour prouver cela, utilisons les densités par rapport à la mesure de comptage ν sur \mathbb{N} . Nous avons :

$$f_{X_n}(k) = \frac{\binom{M}{k} \binom{N-M}{n-k} C_M^k C_{M-N}^{n-k}}{\binom{N}{n}}, 0 \leq k \leq \min(n, M).$$

Supposons que $M/N \rightarrow p$, $N \rightarrow \infty$. Nous aurons

$$\begin{aligned} f_{X_n}(k) &= \frac{M!}{k!(M-k)!} \frac{(M-N)!}{(n-k)!(N-M-(n-k))!} \frac{n!(N-n)!}{N!} \\ &= \frac{n!}{k!(n-k)!} \times \frac{M!}{(M-k)!} \times \frac{!(N-M-(n-k))!}{(N-M-(n-k))!} \times \frac{(M-n)!}{N!} \\ &= \binom{n}{k} \times \left\{ \frac{M!}{(M-k)!} \right\} \left\{ \frac{(N-M)!}{(N-M-(n-k))!} \right\} \left\{ \frac{(M-n)!}{N!} \right\}. \end{aligned}$$

Mais

$$\begin{aligned} \left\{ \frac{M!}{(M-k)!} \right\} &= (M-k+1)(M-k+2)\dots(M+1)M \\ &= M^k \left(1 - \frac{k-1}{M}\right) \left(1 - \frac{k-2}{M}\right) \times \dots \times \left(1 - \frac{1}{M}\right) \\ &= M^k(1 + o(1)) \end{aligned}$$

puisque $M \rightarrow \infty$ et k est fixe. Ensuite

$$\begin{aligned} \left\{ \frac{(N-M)!}{(N-M-(n-k))!} \right\} &= (N-M-(n-k)+1) \times \dots \times (N-M-1)(N-M) \\ &= (N-M)^{n-k} \left(1 + \frac{n-k-1}{N-M}\right) \left(1 + \frac{n-k-2}{N-M}\right) \times \dots \times \left(1 + \frac{1}{N-M}\right) \\ &= (N-M)^{n-k} - 1 + o(1) \end{aligned}$$

puisque, aussi, $N-M = N(1-M/N) \sim N(1-p) \rightarrow \infty$ et $n-k$ est fixé. Enfin

$$\begin{aligned} \left\{ \frac{(M-n)!}{N!} \right\} &= \frac{1}{(N-n+1)(N-n+2)\dots(N-1)N} \\ &= \frac{1}{N^n \left(1 - \frac{n-1}{N}\right) \left(1 - \frac{n-2}{N}\right) \dots \left(1 - \frac{1}{N}\right)} \\ &= \frac{1}{N^n(1 + o(1))}. \end{aligned}$$

pour des raisons similaires. Au total, pour tout $0 \leq k \leq n$

$$f_{X_n}(k) = \binom{n}{k} \left(\frac{M}{N}\right)^k \left(\frac{N-M}{N}\right)^{n-k} (1 + o(1)) \rightarrow \binom{n}{k} p^k (1-p)^{n-k}.$$

Ainsi pour tout point k du domaine de la densité de la loi binomiale par rapport à la mesure de comptage ν , notée

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

nous avons

$$\forall (1 \leq k \leq n), f_{X_n}(k) \rightarrow f_X(k).$$

Nous avons donc la convergence en loi cherchée.

Remarque utile en théorie des sondages. Cette approximation permet de considérer que le tirage avec remise et celui sans remise sont équivalente dans une enquête avec une population très large. C'est un peu la notion suivante : Lorsque la population est très grande, dans un tirage avec remise d'un nombre d'individu assez modeste, il est presque

impossible qu'un même individu sorte plus d'une fois.

3.2. Convergence de la loi binomiale vers la loi de Poisson.

Soit X_n suivant une loi $\mathcal{B}(n, p)$ aec $p = p_n \rightarrow 0$ et $np_n \rightarrow \lambda$, $0 < \lambda$, quand $n \rightarrow \infty$. Alors X_n tend en loi vers une variable aléatoire X suivant une loi de Poisson de paramètre λ .

PREUVE. Pour attester cela, utilisons les fonctions des moments. Soit X une variable suivant la loi de Poisson de paramètre $\lambda > 0$. Nous avons

$$\Psi_{X_n}(t) = (p_n + (1 - p_n)e^t)^n \text{ pour } n \geq 1; \Psi_X(t) = \exp(\lambda(e^t - 1)), t \in \mathbb{R}.$$

Notons $\lambda_n = np_n \rightarrow \lambda$. Pour tout t fixé,

$$\Psi_{X_n}(t) = \left(\frac{\lambda_n}{n} + \left(1 - \frac{\lambda_n}{n}\right)e^t\right)^n = \left(1 - \frac{\lambda_n(e^t - 1)}{n}\right)^n \rightarrow \exp(\lambda(e^t - 1)) = \Psi_X(t)$$

par le résultat classique d'analyse qui affirme que

$$\left(1 + \frac{x_n}{n}\right)^n \rightarrow e^x \text{ quand } n \rightarrow +\infty \text{ pourvu que } x_n \rightarrow x \in \mathbb{R} \text{ lorsque } n \rightarrow +\infty.$$

3.3. Convergence de la loi de Poisson vers la loi normale.

Soit Z_λ une variable aléatoire suivant la loi de Poisson de paramètre λ : $Z_\lambda \sim \mathcal{P}(\lambda)$. Alors la variable

$$\frac{Z_\lambda - \lambda}{\sqrt{\lambda}}$$

converge en loi vers une variable aléatoire X suivant une loi normale standard, i.e $X \sim \mathcal{N}(0, 1)$, quand $\lambda \rightarrow +\infty$.

PREUVE. Nous utilisons les fonctions génératrices des moments. Rappelons la fonction des moments de $Z_\lambda \sim \mathcal{P}(\lambda)$:

$$\Psi_{Z_\lambda}(t) = \exp(\lambda(e^t - 1)).$$

Soit

$$Y(\lambda) = \frac{Z - \lambda}{\sqrt{\lambda}} = \frac{Z - \mathbb{E}(X)}{\sigma_Z}.$$

Nous avons

$$\Psi_{Y(\lambda)}(u) = e^{-\sqrt{\lambda}} \times \varphi_Z(u/\sqrt{\lambda}) = e^{-\sqrt{\lambda}} \times \exp(\lambda(e^{u/\sqrt{\lambda}} - 1)).$$

Quand $\lambda \rightarrow \infty$, nous pouvons développer

$$\begin{aligned}\lambda(e^{u/\sqrt{\lambda}} - 1) &= \lambda\left(1 + \frac{u}{\sqrt{\lambda}} + \frac{u^2}{2\lambda} + O(\lambda^{-3/2}) - 1\right) \\ &= u\sqrt{\lambda} + \frac{u^2}{2} + O(\lambda^{-1/2}).\end{aligned}$$

Donc

$$\Psi_{Y(\lambda)}(u) = \exp\left(\frac{u^2}{2} + O(\lambda^{-1/2})\right) \rightarrow \exp(u^2/2).$$

On conclut aussi que

$$\frac{Z - \lambda}{\sqrt{\lambda}} \rightarrow \mathcal{N}(0, 1)$$

quand $\lambda \rightarrow \infty$.

3.4. Convergence de la loi binomiale vers la loi normale.

Soit X_n suivant une loi $\mathcal{B}(n, p)$ avec $p \in]0, 1[$ fixé. Alors quand $n \rightarrow \infty$

$$(3.1) \quad Z_n = \frac{X_n - np}{\sqrt{npq}} \rightsquigarrow N(0, 1) \quad \text{a} \quad \text{asn} \rightarrow +\infty.$$

Preuve. Utilisons les fonctions des moments. Soit $X \sim \mathcal{B}(n, p)$. Nous avons

$$\Psi_{X_n}(u) = (q + pe^u)^n.$$

D'où

$$(3.2) \quad \Psi_{(X_n - np)/\sqrt{npq}}(u) = e^{-\sqrt{np/q}} \times \Psi_{X_n}(u/\sqrt{npq})$$

avec

$$\Psi_X(u/\sqrt{npq}) = (q + pe^{u/\sqrt{npq}})^n.$$

L'idée de la suite des calculs est d'utiliser un développement d'ordre 2 de $e^{u/\sqrt{npq}}$ au voisinage de 0 quand $n \rightarrow \infty$ et u fixé. Ensuite l'expression obtenue sera de la forme $1 + v_n$, où v_n tend vers zéro. Ensuite, un développement de $\log(1 + v_n)$ d'ordre 2 est opéré.

Ainsi, quand $n \rightarrow \infty$ et u fixé,

$$e^{u/\sqrt{npq}} = 1 + \frac{u}{\sqrt{npq}} + \frac{u^2}{2npq} + O(n^{-3/2}).$$

D'où

$$(q + pe^{u/\sqrt{npq}}) = 1 + u\sqrt{p/nq} + \frac{u^2}{2nq} + O(n^{-3/2}) = 1 + v_n$$

avec

$$v_n = u\sqrt{p/nq} + \frac{u^2}{2nq} + O(n^{-3/2}) \rightarrow 0.$$

Si bien que

$$\begin{aligned} \log(1 + u\sqrt{p/nq} + \frac{u^2}{2nq} + O(n^{-3/2})) &= \log(1 + v_n) \\ &= v_n - \frac{1}{2}v_n^2 + O(v_n^3) \\ &= u\sqrt{p/nq} + \frac{u^2}{2nq} - \frac{pu^2}{2nq} + O(n^{-3/2}). \end{aligned}$$

Alors

$$\begin{aligned} \Psi_{X_n}(u/\sqrt{npq}) &= (q + pe^{u/\sqrt{npq}})^n = \exp(n \log(q + pe^{u/\sqrt{npq}})) \\ &= \exp(n(u\sqrt{p/nq} + \frac{u^2}{2nq} - \frac{pu^2}{2nq} + O(n^{-3/2}))) \\ &= \exp(u\sqrt{np/q} + \frac{u^2}{2q} - \frac{pu^2}{2q} + O(n^{-1/2})) \\ &= e^{u\sqrt{np/q}} e^{u^2/2 + O(n^{-1/2})}. \end{aligned}$$

En retournant à (3.2), on arrive à

$$\Psi_{(X_n - np)/\sqrt{npq}}(u) \rightarrow \exp(u^2/2).$$

D'où l'approximation

$$(\beta(n, p) - np)/\sqrt{npq} \rightarrow \mathcal{N}(0, 1)$$

QED.

Remarque. Nous reviendrons sur une deuxième preuve directe de ce résultat en utilisant le théorème central limite standard ci-dessous.

3.5. Théorème Central Limite Standard dans \mathbb{R} . Les deux cas déjà vus sont des cas spéciaux d'un cas général. En effet, si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires réelles ayant des moments de second ordre finis, on peut s'attendre à ce que

$$\frac{X_n - E(X_n)}{\sigma_{X_n}}$$

converge en loi vers la loi normale centrée réduite. Ceci n'est pas toujours vraie. Mais si cela est le cas, nous dirons qu'on a un théorème central limite (Central limit théorème, CLT). Cela est vrai pour l'échantillon dans le cas suivant.

Soit X_1, X_2, \dots une suite de variables aléatoires indépendantes, identiquement distribuées selon la fonction de répartition F avec

$$E(X_i) = \mu = \int x dF(x) = 0, \sigma_{X_i}^2 = \sigma^2 = \int (x - \mu)^2 dF(x).$$

Posons, pour $n \geq 1$,

$$S_n = X_1 + \dots + X_n.$$

Alors quand $n \rightarrow \infty$

$$\frac{S_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1).$$

PREUVE. Considérons les fonctions caractéristiques

$$\mathbb{R} \ni u \mapsto \Phi_{X_i}(u) = E(e^{iuX_i}) = \Psi(u).$$

Par l'existence des moments à l'ordre 2, on a le développement limité à l'ordre 2,

$$\begin{aligned} \Phi(u) &= 1 + iu\Phi'(0) + \frac{1}{2}u^2\Phi''(0) + O(u^3) \\ &= 1 - \frac{1}{2}u^2 + O(u^2) \end{aligned}$$

puisque

$$\Phi'(0) = i \mathbb{E}(X) = 0, \quad \Phi''(0) = -\mathbb{E}(X^2) = -1.$$

Dès lors

$$\Phi_{S_n/\sqrt{n}}(u) = (\Phi(u/\sqrt{n}))^n.$$

Pour u fixé et $n \rightarrow \infty$,

$$\begin{aligned} \Phi_{S_n/\sqrt{n}}(u) &= (\Phi(u/\sqrt{n}))^n = \exp(n \log(1 - \frac{u^2}{2n} + O(n^{-3/2}))) \\ &= \exp(n(-\frac{u^2}{2n} + O(n^{-3/2}))) \\ &= \exp(-u^2/2 + O(n^{-1/2})) \\ &\rightarrow \exp(-u^2/2) \end{aligned}$$

Nous venons d'établir que

$$\frac{S_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1).$$

Dans le cas général non centré et non normalisé, nous avons le Théorème central limite

$$\frac{1}{\sigma\sqrt{n}}(S_n - n\mu) \rightarrow \mathcal{N}(0, 1).$$

Donnons deux exemples importants du théorème central limite standard relatifs aux expériences de Bernouilli.

Exemple 1 : Convergence vague de la loi binomiale.

Nous allons donner une autre preuve du résultat (3.1) de la sous-section 3.4 relatif la loi limite d'une suite de variables aléatoires suivant une loi binomiale lorsque le nombre d'essais n croît indéfiniment pendant que la probabilité de succès $\in]0, 1[$ reste fixé. Nous gardons les notations de cette sous-section.

En nous fondant sur les cours de probabilités élémentaires que nous pouvons trouver dans un grand nombre d'ouvrages, en particulier dans [5], de la présente série de Théorie de Probabilité et de Statistiques, au chapitre, Lemme 1, que si $X_n \sim \mathcal{B}(n, p)$, alors X_n est la somme de n variables aléatoires Y_1, \dots, Y_n , identiquement distribuées selon un loi de Bernouilli $\mathcal{B}(p)$ random variables, i.e.,

$$X_n = Y_1 + \dots + Y_n.$$

Pour chaque Y_i , $1 \leq i \leq n$, nous avons

$$\mathbb{E}(Y_i) = p \text{ and } \sigma^2 = \mathbb{V}ar(Y_i) = pq \text{ where } q = 1 - p.$$

Dès lors, la variable Z_n de la formule (3.1) devient

$$Z_n = \frac{X_n - np}{\sqrt{npq}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)).$$

Ainsi, la convergence vague de Z_n to $\mathcal{N}(0, 1)$ quand $n \rightarrow +\infty$ découle de l'application du théorème central limite sur \mathbb{R} .

Remarque. Cette preuve est simple et belle. La première est toujours utile. Simplement parce que nous pouvons tre appelés utiliser ou à enseigner ce résultat à un niveau où le théorème central limite n'est pas disponible. Au delà de cette raison, cette preuve fait partie de l'histoire des probabilités. Dans le même esprit, les premières découvertes de cette loi remontent en 1732 avec *de Moivre* et en 1801

avec Laplace (voir Loève [9], page 23). Ces méthodes historiques peuvent être revisités dans [5] ou dans [6] avec une rédaction adaptée au niveau de la première année universitaire.

Exemple 2 : Loi Binomiale Négative.

Pour un entier $k \geq 1$ fixée, une variable suivant la loi Binomiale Négative X_k peut être définie relativement aux essais de Bernoulli avec une probabilité de succès $p \in]0, 1[$. Le nombre d'essais indépendants de l'expérience de Bernoulli de p nécessaires pour avoir k succès, suit par définition la loi Binomiale Négative de paramètres k et p , notée $X_k \sim \mathcal{NB}(k, p)$. Pour $k = 1$, la X_1 suit une loi géométrique de paramètre p , notée $X_1 \sim \mathcal{G}(p)$.

De même que pour la suite de variables aléatoires binomiales, nous pouvons appliquer le une suite de variables aléatoires binomiales négatives X_k , $k \geq 1$ pour avoir le résultat

$$(3.3) \quad Z_n = \frac{p(X_k - \frac{k}{p})}{\sqrt{npq}} \rightsquigarrow \mathcal{N}(0, 1) \text{ quand } k \rightarrow +\infty.$$

A cet effet, le lecteur peut trouver dans les cours de probabilités élémentaires de son choix, en particulier dans [5], de la présente série de Théorie de Probabilité et de Statistiques, au chapitre 2, que le lemme 2 assure que la variable aléatoire X_k suivant la loi $\mathcal{NB}(k, p)$ est la somme de k variables aléatoires Y_1, \dots, Y_k , indépendantes et identiquement distribuées selon la loi géométrique $\mathcal{G}(p)$, i.e.,

$$X_k = Y_1 + \dots + Y_n,$$

et que pour chacune des variables Y_i , nous avons

$$\mathbb{E}(Y_i) = \frac{1}{p} \text{ et } \sigma^2 = \text{Var}(Y_i) = \frac{q}{p^2} \text{ où } q = 1 - p.$$

Alors, en appliquant le théorème central limite, nous obtenons

$$Z_n = \frac{p(X_k - \frac{k}{p})}{\sqrt{npq}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)) \rightsquigarrow \mathcal{N}(0, 1) \text{ as } k \rightarrow +\infty.$$

Ceci prouve (3.3).

3.6. Lois limites des valeurs extrêmes. Considérons X_1, X_2 , etc..., une suite de variables aléatoires réelles indépendantes de même fonction de répartition F . Considérons pour chaque $n \geq 1$

$$M_n = \max(X_1, \dots, X_n).$$

Rappelons tout de suite que nous avons

$$P(M_n \leq x) = F(x)^n, x \in \mathbb{R}.$$

La théorie des valeurs extrêmes a commencé ses beaux jours par la découverte des lois limites de la suite M_n en type. On dira que M_n converge en type vers Z si et seulement il existe des suites $(a_n > 0)_{n \geq 1}$ et $(b_n)_{n \geq 1}$ telles que la suite de variables aléatoires

$$\frac{M_n - b_n}{a_n}$$

converge vaguement vers Z ,

$$\frac{M_n - b_n}{a_n} \rightsquigarrow Z.$$

Découvrons les trois limites trois types de lois extrémales définies par leur fonctions de répartition ($f.r$):

Type de Gumbel : $\Lambda(x) = \exp(-\exp(-x)), x \in \mathbb{R}$. (On note Λ une $v.a$ de $f.r$ Λ)

Type de Frechet de paramètre $\alpha > 0$:

$$\varphi_\alpha(x) = \exp(-x^{-\alpha})1_{(x \geq 0)}.$$

On note $FR(\alpha)$ une $v.a$ de $f.r$ φ_α)

Type de Weibull de paramètre $\beta > 0$:

$$\psi_\beta(x) = \exp(-(-x)^\beta)1_{(x < 0)} + 1_{(x \geq 0)}.$$

(On note $W(\beta)$ une $v.a$ de $f.r$ ψ_β).

Donnons quelques exemples de convergence en type de maxima.

Nous allons utiliser la convergence des fonction de répartition. Puisque les limites vagues ont des fonction de répartition continues, nous allons

voir la limite en tout point de \mathbb{R} .

(a) Loi de exponentielle : Supposons que

$$F(x) = (1 - \exp(-x))1_{(x \geq 0)}$$

est celle d'une loi exponentielle standard. En utilisant les fonctions de répartition, montrons que

$$M_n - \log n \xrightarrow{d} \Lambda.$$

En effet

$$P(M_n - \log n \leq x) = P(M_n \leq x + \log n) = F(M_n \leq x + \log n)^n.$$

Pour tout $x \in \mathbb{R}$, $x + \log n \geq 0$ pour $n \geq \exp(-x)$. Donc pour n assez grand $F(M_n \leq x + \log n) = (1 - \exp(-x - \log n))$ et donc pour tout $x \in \mathbb{R}$

$$P(M_n - \log n \leq x) = (1 - \frac{e^{-x}}{n}) \rightarrow e^{-e^{-x}} = \Lambda(x).$$

(b) Loi de paréto de paramètre $\alpha > 0$:

$$F(x) = (1 - x^{-\alpha})1_{(x \geq 1)}.$$

En utilisant les fonctions de répartition, montrons que

$$n^{-1/\alpha} M_n \xrightarrow{d} C(\alpha).$$

Remarquons que les variables de Pareto sont positives et donc M_n est positif pour tout $n \geq 1$. Etudions deux cas.

Cas $x \leq 0$. Dans ce cas

$$P(n^{-1/\alpha} M_n \leq 0) = 0 = \varphi_\alpha(x),$$

et la limite des fonctions de répartition a lieu.

Cas $x > 0$. Dans ce cas

$$P(n^{-1/\alpha} M_n \leq x) = P(M_n \leq n^{1/\alpha} x).$$

Donc pour n assez grand, on aura $n^{1/\alpha} x > 1$ (par exemple prendre $n \geq (1/x)^{-\alpha}$) et pour ces n ,

$$\begin{aligned} P(n^{-1/\alpha} M_n \leq x) &= F(n^{1/\alpha} x)^n = (1 - (n^{1/\alpha} x)^{-\alpha})^n \\ &= (1 - \frac{x^{-\alpha}}{n})^n \rightarrow \exp(-x^{-\alpha}) \\ &= \varphi_\alpha(x). \end{aligned}$$

Nous avons bien que pour tout $x \in \mathbb{R}$,

$$P(n^{-1/\alpha} M_n \leq x) \longrightarrow \varphi_\alpha(x).$$

Dès lors

$$n^{-1/\alpha} M_n \rightsquigarrow FR(\alpha).$$

(c) Loi uniforme sur $(0, 1)$:

$$F(x) = x1_{(0 \leq x \leq 1)} + 1_{(x \geq 1)}.$$

En utilisant les fonctions de répartition, montrons que

$$n(M_n - 1) \xrightarrow{d} W(1).$$

Nous avons

$$P(n(M_n - 1) \leq x) = F(1 + \frac{x}{n})^n.$$

Etudions deux cas.

Cas $x \geq 0$. Dans ce cas $1 + x/n$ est positif pour tout $n \geq 1$ et

$$P(n(M_n - 1) \leq x) = F(1 + \frac{x}{n})^n = 1 = \psi_1(x)$$

et nous avons la convergence des fonctions de répartition.

Cas $x < 0$. Pour n suffisamment grand, nous aurons $0 \leq 1 + x/n \leq 1$ (prendre par exemple $x \geq -n$ i.e. $n \geq -(x) \geq 0$). Pour ces valeurs de n ,

$$\begin{aligned} P(n(M_n - 1) \leq x) &= F(1 + \frac{x}{n})^n \\ &= (1 + \frac{x}{n})^n \rightarrow e^x = \psi_1(x). \end{aligned}$$

Nous avons bien que pour tout $x \in \mathbb{R}$,

$$P(n(M_n - 1) \leq x) \longrightarrow \psi_1(x).$$

Dès lors

$$n(M_n - 1) \rightsquigarrow W(1).$$

Résumé : En théorie des valeurs extrêmes, il est démontré que les seules lois limites non dégénérées possibles sont bien celles-la. Dans le chapitre réservé à une étude spécifique de la convergence dans \mathbb{R} utilisant les inverses généralisées, les formes générales des distributions dont les maxima convergent chacun des types seront données.

4. Exemples de convergence dans \mathbb{R}^k

4.1. Théorème Central Limite Standard sur \mathbb{R}^k . Passons au théorème central limite dans \mathbb{R}^k . Soit X_1, X_2, \dots une suite de variables aléatoires centrées, indépendantes et identiquement distribuées (i.i.d) centrées de matrice de covariance $\Sigma = (\sigma_{ij})_{1 \leq i \leq k, 1 \leq j \leq k}$, c'est à dire

$$\sigma_{ij} = Cov(X_i, X_j).$$

Considérons les sommes partielles

$$S_n = X_1 + X_2 + \dots + X_n.$$

Nous avons le théorème central limit sur \mathbb{R}^k ,

$$S_n/\sqrt{n} \rightsquigarrow \mathcal{N}(0, \Sigma)$$

PREUVE. Remarquons que la matrice Σ est symétrique et semi-positive puisque, pour tout $u \in \mathbb{R}^k$

$${}^t u \Sigma u = {}^t u \mathbb{E}(X X') u = \mathbb{E}(({}^t X u)({}^t X u)) = \mathbb{E}(({}^t X u)^2) \geq 0.$$

D'après la théorie des matrices (voir cours d'algèbre de deuxième année), Σ possède des valeurs propres non négatives et elle est diagonalisable au moyen d'une matrice orthogonale T , telle que

$${}^t T \Sigma T = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \Lambda.$$

Posons

$$Y_i = {}^t T X_i.$$

Les variables Y_i sont centrées, iid et de matrice de covariance

$$\Sigma_Y = {}^t T \Sigma T = \Lambda.$$

Cela veut dire que les composantes de Y_i sont non corréolées et ont pour variances respectives les nombres $\lambda_1, \lambda_2, \dots, \lambda_n$. Posons

$$(4.1) \quad M_n = \frac{1}{\sqrt{n}}(Y_1 + Y_2 + \dots + Y_n) = {}^t T \left(\frac{S_n}{\sqrt{n}} \right).$$

Pour tout $A = {}^t(a_1, a_2, \dots, a_k) \in \mathbb{R}^k$,

$$\langle A, M_n \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{i=n} \langle A, Y_i \rangle.$$

Or les variables $\langle A, Y_i \rangle$ sont centrée, i.i.d, de variance

$$\mathbb{E} \langle A, Y_i \rangle^2 = \sum_{i=1}^{i=n} a_i^2 \lambda_i = {}^t A \Lambda A,$$

en vertu de la non corrélation des composantes de chaque Y_i . Le théorème central limite dans \mathbb{R} implique que

$$\langle A, M_n \rangle \rightarrow \mathcal{N}(0, \sum_{i=1}^{i=n} a_i^2 \lambda_i) = \mathcal{N}(0, {}^t A \Lambda A)$$

Or $\mathcal{N}(0, {}^t A \Lambda A)$ est la loi d'un vecteur gaussien résultant de la transformation linéaire ${}^t A Z = \langle A, Z \rangle$, où Z suit la loi $\mathcal{N}(0, \Lambda)$. D'où

$$\forall A \in \mathbb{R}^k, \langle A, M_n \rangle \rightsquigarrow \langle A, Z \rangle.$$

En terme de fonction caractéristique, cela veut dire que pour $t \in \mathbb{R}$ et pour tout $A \in \mathbb{R}^k$,

$$\mathbb{E} \exp(it \langle A, M_n \rangle) \rightarrow \mathbb{E} \exp(it \langle A, Z \rangle).$$

Pour $t = 1$, nous avons pour tout $A \in \mathbb{R}^k$

$$\Phi_{M_n}(A) = \mathbb{E} \exp(it \langle A, M_n \rangle) \rightarrow \Phi_Z(A) = \mathbb{E} \exp(it \langle A, Z \rangle).$$

Ceci veut bien dire que

$$M_n \rightsquigarrow Z.$$

Ceci, la formule (4.1) et le point (e) du théorème 1 ensemble impliquent

$$S_n / \sqrt{n} = T M_n \rightarrow T Z$$

et

$${}^t T Z \sim \mathcal{N}(0, T \Lambda {}^t T) = \mathcal{N}(0, \Sigma).$$

D'où, enfin,

$$S_n / \sqrt{n} \rightsquigarrow \mathcal{N}(0, \Sigma)$$

Ceci exprime la version simple du théorème central limite dans \mathbb{R}^k .

4.2. Convergence de la loi multinomiale. Un k -uplet $X_n = (X_{1,n}, \dots, X_{k,n})$ suit une loi multinomiale de paramètres $n \geq 1$ et $p = (p_1, p_2, \dots, p_k)$ avec

$$\forall (1 \leq i \leq k), p_i > 0 \text{ et } \sum_{1 \leq i \leq k} p_i = 1,$$

notée $\mathcal{M}_k(n, p)$, ssi sa loi de probabilité est:

$$\mathbb{P}(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1! \times \dots \times n_k!} p_1^{n_1} \times p_2^{n_2} \times \dots \times p_k^{n_k}$$

pour (n_1, \dots, n_k) vérifiant

$$\forall (1 \leq i \leq k), n_i \geq 0 \text{ et } \sum_{1 \leq i \leq k} n_i = n.$$

Elle est générée de la manière suivante. Soit une expérience à k issues $E_i, 1 \leq i \leq k$, chacune se réalisant avec une probabilité $p_i > 0$. On la répète n fois de manière indépendante. A l'issue de ces n essais, soit $X_{i,n}$ le nombre de réalisations de l'issue E_i . Le vecteur ainsi obtenu suit une loi $\mathcal{M}_k(n, p)$. Bien sûr chaque $X_{i,n}$ suit une loi binomiale $\mathcal{B}(n, p_i)$.

Nous avons le résultat de convergence vague suivant.

$$(4.2) \quad Z_n = {}^t \left(\frac{X_{1,n} - np_1}{\sqrt{np_1}}, \dots, \frac{X_{k,n} - np_k}{\sqrt{np_k}} \right) \rightsquigarrow \mathcal{N}_k(0, \Sigma) \text{ as } n \rightarrow +\infty,$$

où Σ est matrice carrée d'ordre k dont les éléments sont $\Sigma_{i,i} = 1 - p_i$ et $\Sigma_{i,j} = \sqrt{p_i p_j}$, $1 \leq i, j \leq k$.

Remarque importante. Ce résultat a d'importantes applications. Nous pouvons mentionner son utilisation pour trouver la loi des distributions finies du processus empirique en probabilités. Il est aussi la base des tests du khi-deux en Statistiques. Nous verrons ces tests plus tard dans la partie réservée cet effet dans cette série.

Preuve. Nous allons présenter deux preuves. La première utilise la convergence de la fonction des moments et du development de la fonction logarithmique. Elle est plus adaptée un enseignement de premier cycle. La deuxième exploite le théorème central limit dans \mathbb{R}^k , que nous avons vu précédemment. Elle est plus adaptée exposé de niveau supérieur.

Première preuve.

Nous connaissons déjà sa fonction génératrice des moments qui est

$$\phi_{X_n}(u) = \left(\sum_{1 \leq i \leq k} p_i e^{u_i} \right)^n.$$

Posons

$$\begin{aligned} Z_n &= \left(\frac{X_{1,n} - np_1}{\sqrt{np_1}}, \dots, \frac{X_{k,n} - np_k}{\sqrt{np_k}} \right) \\ &= AX + B \end{aligned}$$

avec

$$A = \begin{pmatrix} 1/\sqrt{np_1} & & & \\ & \sqrt{np_2} & & \\ & & \dots & \\ & & & \sqrt{np_k} \end{pmatrix}$$

et

$$B = \begin{pmatrix} -\sqrt{np_1} \\ -\sqrt{np_2} \\ \dots \\ -\sqrt{np_k} \end{pmatrix}.$$

Dès lors

$$\begin{aligned} \phi_{Z_n}(u) &= \exp(\langle B, u \rangle) \times \phi_X({}^t A u) \\ &= \exp\left(\sum_{1 \leq i \leq k} -\sqrt{np_i} u_i\right) \times \left(\sum_{1 \leq i \leq k} p_i e^{u_i/\sqrt{np_i}}\right)^n \end{aligned}$$

Notons que u est fixé. Pour tout i fixé, $u_i/\sqrt{np_i} \rightarrow 0$ quand $n \rightarrow \infty$ car chaque $p_i > 0$. D'où

$$e^{u_i/\sqrt{np_i}} = 1 + u_i/\sqrt{np_i} + \frac{1}{2} \frac{u_i^2}{np_i} + O(n^{-3/2}).$$

D'où

$$\begin{aligned} A &= \left(\sum_{1 \leq i \leq k} p_i e^{u_i/\sqrt{np_i}}\right)^n = \exp\left(n \log\left(\sum_{1 \leq i \leq k} p_i e^{u_i/\sqrt{np_i}}\right)\right). \\ &= \exp\left(n \log\left(1 + \sum_{1 \leq i \leq k} u_i \sqrt{p_i/n} + \sum_{1 \leq i \leq k} \frac{1}{2} \frac{u_i^2}{n} + O(n^{-3/2})\right)\right). \end{aligned}$$

Posons aussi

$$a = \sum_{1 \leq i \leq k} u_i \sqrt{p_i/n} + \sum_{1 \leq i \leq k} \frac{1}{2} \frac{u_i^2}{n} \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Nous aurons

$$A = \exp(n \log(1 + a)).$$

Développons $\log(1 + a)$ à l'ordre 2 en mettant dans $O(n^{-3/2})$ tous les autres termes tendant vers zéro:

$$\begin{aligned}
A &= \exp\left(n\left(a - \frac{1}{2}a^2 + O(a^3)\right)\right) \\
&= \exp\left(n\left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i/n} + \sum_{1 \leq i \leq k} \frac{1}{2} \frac{u_i^2}{n} - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i/n}\right)^2 + O(n^{-3/2})\right)\right) \\
&= \exp\left(\sum_{1 \leq i \leq k} u_i \sqrt{np_i} + \sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2 + O(n^{-1/2})\right) \\
&= \exp\left(\sum_{1 \leq i \leq k} u_i \sqrt{np_i}\right) \times \exp\left(\sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2 + O(n^{-1/2})\right).
\end{aligned}$$

En mettant tout cela ensemble, nous obtenons

$$\begin{aligned}
\phi_{Z_n}(u) &= \exp\left(\sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2 + O(n^{-1/2})\right) \\
&\rightarrow \phi_Z(u) = \exp\left(\sum_{1 \leq i \leq k} \frac{1}{2} u_i^2 - \frac{1}{2} \left(\sum_{1 \leq i \leq k} u_i \sqrt{p_i}\right)^2\right).
\end{aligned}$$

Et

$$(4.3) \quad \phi_Z(u) = \exp\left\{\sum_{1 \leq i \leq k} \frac{1}{2} (1 - p_i) u_i^2 - \sum_{1 \leq i, j \leq k} u_i u_j \sqrt{p_i p_j}\right\}$$

est la fonction des moments d'un vecteur gaussien Z centré dont la matrice de variances-covariances Σ vérifie

$$(4.4) \quad \Sigma_{ii} = (1 - p_i)$$

et

$$(4.5) \quad \Sigma_{ij} = -\sqrt{p_i p_j}.$$

Donc

$$Z_n \rightsquigarrow \mathcal{N}(O, \Sigma).$$

La première preuve finit ici.

Deuxième preuve. Au i -ième, $i \in \{1, \dots, n\}$, nous obtenons le vecteur variable aléatoire

$$Z^{(i)} = \begin{pmatrix} Z_1^{(i)} \\ \dots \\ Z_k^{(i)} \end{pmatrix}$$

défini ainsi : pour chaque $1 \leq r \leq k$, nous avons

$$Z_r^{(i)} = \begin{cases} 1 & \text{si l'issue } E_r \text{ se réalise au } i\text{-ième essai et aucune autre ne se réalise} \\ 0 & \text{si un autre issue que } E_r \text{ se réalise} \end{cases}$$

Il est évident que $Z^{(i)}$ a une distribution multinomiale $\mathcal{M}_k(1, k)$ et que les vecteurs aléatoires $Z^{(i)}$ sont indépendents.

De plus, pour tout $i \in \{1, \dots, n\}$, chaque $Z_r^{(i)}$, $1 \leq r \leq k$, suit une loi de Bernoulli de paramètre p et un seul des $Z_r^{(i)}$ ($1 \leq r \leq k$) prend la valeur un (1), les autres étant nuls. Cela implique que

$$Z_r^{(i)} Z_s^{(i)} = 0 \text{ for } 1 \leq r \neq s \leq k, 1 \leq i \leq n.$$

Nous avons aussi

$$Z_1^{(i)} + \dots + Z_k^{(i)} = 1.$$

Alors, pour tout $i \in \{1, \dots, n\}$,

$$\mathbb{E}(Z_r^{(i)}) = p_i \text{ et } \mathbb{V}ar(Z_r^{(i)}) = p_i(1 - p_i), \quad 1 \leq r \leq k$$

et pour tout $1 \leq r \neq s \leq k$

$$\text{cov}(Z_r^{(i)}, Z_s^{(i)}) = \mathbb{E}(Z_r^{(i)} Z_s^{(i)}) - \mathbb{E}(Z_r^{(i)})\mathbb{E}(Z_s^{(i)}) = -p_r p_s,$$

puisque $Z_r^{(i)} Z_s^{(i)} = 0$. Dès lors, chaque $Z^{(i)}$ possède la matrice de variance-covariance

$$\Sigma^0 = \begin{pmatrix} p_1(1 - p_1) & -p_1 p_2 & \dots & -p_1 p_{k-1} & -p_1 p_k \\ -p_2 p_1 & p_2(1 - p_2) & \dots & -p_2 p_{k-1} & -p_2 p_k \\ \dots & \dots & \dots & \dots & \dots \\ -p_{k-1} p_1 & -p_{k-1} p_2 & \dots & p_{k-1}(1 - p_{k-1}) & -p_{k-1} p_k \\ -p_k p_1 & -p_k p_2 & \dots & -p_k p_{k-1} & -p_k(1 - p_k) \end{pmatrix}$$

ou, par une autre notation,

$$\Sigma_0 = (\sigma_{ij}^0)_{1 \leq i, j \leq k} \text{ with } \sigma_{ij}^0 = \begin{cases} p_i(1 - p_i) & \text{if } i = j \\ -p_i p_j & \text{if } i \neq j \end{cases}.$$

Après n essais, la somme des vecteurs aléatoires $Z^{(1)}, \dots, Z^{(n)}$, qui sont indépendants et suivent identiquement une loi $\mathcal{M}_k(1, k)$, donne X_n , autrement

$$X_n = Z^{(1)} + \dots + Z^{(n)}.$$

Par le théorème central limit standard multivarié, nous avons, quand $n \rightarrow +\infty$,

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z^{(i)} - \mathbb{E}(Z^{(i)})) \rightsquigarrow Z_0 \sim \mathcal{N}_k(0, \Sigma_0).$$

Mais nous pouvons aisément vérifier que

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z^{(i)} - \mathbb{E}(Z^{(i)})) = {}^t \left(\frac{X_{1,n} - np_1}{\sqrt{n}}, \frac{X_{2,n} - np_2}{\sqrt{n}}, \dots, \frac{X_{k,n} - np_k}{\sqrt{n}} \right).$$

Et nous obtenons la relation matricielle

$$DS_n = Z_n,$$

où D est la matrice diagonale

$$D = \text{diag}(1/\sqrt{p_1}, \dots, 1/\sqrt{p_k}).$$

Par the théorème de transformation continue (Point (e) du théorème 1), nous avons

$$Z_n = DS_n \rightsquigarrow DZ_0 \sim \mathcal{N}_k(0, D\Sigma_0 D),$$

puisque D est symétrique. Il reste à calculer

$$\Sigma = D\Sigma_0 D = (\sigma_{ij})_{1 \leq i, j \leq k}.$$

For $1 \leq h, j \leq k$, $(\Sigma_0 D)_{hj}$ est la produit matriciel de la h -ième ligne de Σ_0 par la j -ième colonne de D . En exploitant la diagonalité de D , nous obtenons pour $1 \leq h, j \leq k$,

$$(\Sigma_0 D)_{hj} = \sigma_{hj}^0 / \sqrt{p_j}.$$

Ensuite, $\sigma_{ij} = (D\Sigma_0 D)_{ij}$ est le produit de la i -ième ligne de D par la j -ième colonne de $(\Sigma_0 D)^{(j)} = {}^t((\Sigma_0 D)_{1j}, (\Sigma_0 D)_{2j}, \dots, (\Sigma_0 D)_{kj})$. En utilisant encore le fait que D est diagonale, nous arrivons à

$$(D\Sigma_0 D)_{ij} = \frac{1}{\sqrt{p_i}} (\Sigma_0 D)_{ij},$$

ce qui aboutit à

$$\sigma_{ij} = (D\Sigma_0 D)_{ij} = \frac{1}{\sqrt{p_i p_j}} \sigma_{ij}^0 = \begin{cases} \sigma_{ii}^0 / p_i = 1 - p_i & \text{if } i = j \\ -\sqrt{p_i p_j} & \text{if } i \neq j \end{cases} \quad \dots$$

Nous obtenons encore

$$Z_n \rightsquigarrow \mathcal{N}_k(0, \Sigma), \quad \text{as } n \rightarrow +\infty.$$

où Σ is défini dans la ligne qui suit la formule (4.2) dans le partie haut de cette sous-section. Ceci finit la deuxième preuve.

En conclusion, nous avons le résultat suivant.

PROPOSITION 2. Soit une suite de vecteurs aléatoires $X(n) = (X_1(n), \dots, X_k(n))$ suivant une loi multinomiale de paramètres $n \geq 1$ et $p = (p_1, p_2, \dots, p_k)$ avec

$$\forall (1 \leq i \leq k), p_i > 0 \text{ et } \sum_{1 \leq i \leq k} p_i = 1.$$

Alors la suite de vecteurs

$$Z_n = \left(\frac{X_1 - np_1}{\sqrt{np_1}}, \dots, \frac{X_k - np_k}{\sqrt{np_k}} \right)$$

converge vers une loi normale k -dimensionnée, centrée, de matrice de variances-covariances Σ avec

$$\Sigma_{ii} = (1 - p_i)$$

et

$$\Sigma_{ij} = -\sqrt{p_i p_j}.$$

4.3. Limites des dimensions finies du processus empirique uniforme. Considérons U_1, U_2, \dots une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi uniforme sur $(0,1)$, de fonction de répartition commune $F(s) = s1_{(0 \leq s \leq 1)} + 1_{(s \geq 1)}$. Pour $n \geq 1$, on définit la fonction répartition empirique uniforme basée sur l'échantillon U_1, U_2, \dots, U_n la fonction

$$\mathbb{R} \ni x \mapsto U_n(s) = \frac{1}{n} \text{Card}\{i, 1 \leq i \leq n, U_i \leq s\}$$

Et définissons le processus empirique uniforme

$$\alpha_n(s) = \sqrt{n}(U_n(s) - s), 0 \leq s \leq 1.$$

Considérons $0 = t_0 < t_1 < \dots < t_k < t_{k+1} = 1$ et posons

$$Y_n = (\alpha_n(t_1), \dots, \alpha_n(t_{k+1})).$$

Alors nous avons le résultat :

PROPOSITION 3. Les distributions finies du processus empirique uniformes de la forme $(\alpha_n(t_1), \dots, \alpha_n(t_{k+1}))$, avec $0 = t_0 < t_1 < \dots < t_k < t_{k+1} = 1$, vérifient

$$(\alpha_n(t_1), \dots, \alpha_n(t_{k+1})) \rightarrow \mathcal{N}_k(0, (\min(t_i, t_j) - t_i t_j)_{1 \leq i, j \leq k+1})$$

PREUVE. Posons

$$(4.6) \quad Z_n = \left(\frac{\alpha_n(t_1)}{\sqrt{t_1}}, \frac{\alpha_n(t_2) - \alpha_n(t_1)}{\sqrt{t_2 - t_1}}, \dots, \frac{\alpha_n(t_{k+1}) - \alpha_n(t_k)}{\sqrt{t_{k+1} - t_k}} \right).$$

Remarquons que

$$N_n = (nF_n(t_1), nF_n(t_2) - nF_n(t_1), \dots, nF_n(t_{k+1}) - nF_n(t_k))$$

suit une loi multinomiale dont les composantes du vecteur des probabilités des issues sont : $t_1, t_2 - t_1, \dots, t_{k+1} - t_k$. En effet,

$$nF_n(t_j) - nF_n(t_{j-1})$$

est le nombre d'observations tombant dans $]t_{j-1}, t_j]$ et pour tout j , la probabilité qu'une observation tombe dans $]t_{j-1}, t_j]$ est bien $p_j = t_j - t_{j-1}$.

Nous pouvons appliquer le théorème de convergence faible de la loi multinomiale établie dans la sous-section 4.2.

Définissons Z_n en centrant chaque j -ième composante de N_n par $n(t_j - t_{j-1})$ en la normalisant par $\sqrt{n(t_j - t_{j-1})}$.

Rappelons nous que nous avons : $Y_n^T = (\alpha_n(t_1), \dots, \alpha_n(t_{k+1}))$. Nous obtenons la relation matricielle

$$Z_n = AY_n \Leftrightarrow Y_n = BZ_n$$

où la relation $y = Bz$ traduit la correspondance suivante :

$$y_i = \sqrt{t_1}x_1 + \sqrt{t_2 - t_1}x_2 + \dots + \sqrt{t_i - t_{i-1}}x_i, \quad 1 \leq i \leq k+1.$$

D'après la convergence de la loi multinomiale sus-mentionnée, Z_n converge faiblement vers un vecteur gaussien $Z = (Z_1, Z_2, \dots, Z_{k+1})$ vérifiant

$$\mathbb{E}(Z_j^2) = 1 - (t_j - t_{j-1})$$

et

$$\mathbb{E}(Z_i Z_j) = -\sqrt{(t_i - t_{i-1})(t_j - t_{j-1})}.$$

Par le théorème de la transformation continue, (Point (e) du théorème 1), $Y_n = BZ_n$ converge vaguement vers $Y = BZ$, avec

$$Y_i = \sqrt{t_1}Z_1 + \sqrt{t_2 - t_1}Z_2 + \dots + \sqrt{t_i - t_{i-1}}Z_i, \quad 1 \leq i \leq k+1.$$

Soit le vecteur $T = (T_1, \dots, T_{k+1})$ défini par $(t_j - t_{j-1})Z_j = T_j$, $1 \leq i \leq k$, i.e.,

$$Z = \left(\frac{T_1}{\sqrt{t_1}}, \frac{T_2}{\sqrt{(t_2 - t_1)}}, \dots, \frac{T_j}{\sqrt{(t_j - t_{j-1})}}, \dots, \frac{T_{k+1}}{\sqrt{(t_{k+1} - t_k)}} \right)$$

Nous avons

$$\mathbb{E}(T_j^2) = \mathbb{E}((Z_j \sqrt{(t_j - t_{j-1})})^2) = (t_j - t_{j-1})(1 - (t_j - t_{j-1})).$$

et

$$\mathbb{E}(T_i T_j) = \sqrt{(t_j - t_{j-1})(t_i - t_{i-1})} \mathbb{E}(Z_i Z_j) = -(t_j - t_{j-1})(t_i - t_{i-1}).$$

Avant de calculer la covariance de Y_i et de Y_j , vérifions que pour $t_i \leq t_j$, nous avons

$$\begin{aligned} t_i t_j &= \left(\sum_{h=1}^{h=i} (t_h - t_{h-1}) \right) \left(\sum_{r=1}^{r=j} (t_r - t_{r-1}) \right) \\ &= \left(\sum_{h=1}^{h=i} (t_h - t_{h-1}) \right) \left(\sum_{r=1}^{r=i} (t_r - t_{r-1}) + \sum_{r=i+1}^{r=j} (t_r - t_{r-1}) \right) \\ &= \left(\sum_{h=1}^{h=i} (t_h - t_{h-1}) \right)^2 + \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} (t_h - t_{h-1})(t_r - t_{r-1}) \\ &= \sum_{h=1}^{h=i} (t_h - t_{h-1})^2 + \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) \\ &\quad - \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} (t_h - t_{h-1})(t_r - t_{r-1}) \end{aligned}$$

En mettant ensemble les points précédents, nous sommes en mesure de calculer la matrice de variance-covariance de Y . Pour $1 \leq i \leq j \leq 1$, nous avons

$$\begin{aligned} Y_i Y_j &= \left(\sum_{h=1}^{h=i} T_h \right)^2 + \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} T_h T_r \\ &= \sum_{h=1}^{h=i} T_h^2 + \sum_{1 \leq h \neq r \leq i} T_h T_r + \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} T_h T_r. \end{aligned}$$

Finalement, nous obtenons

$$\begin{aligned}
\mathbb{E}(Y_i Y_j) &= \sum_{h=1}^{h=i} (1 - (t_h - t_{h-1})) - \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&- \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&= \sum_{h=1}^{h=i} (t_h - t_{h-1}) - \sum_{h=1}^{h=i} (t_h - t_{h-1})^2 \\
&- \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) - \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&= t_i - \sum_{h=1}^{h=i} (t_h - t_{h-1})^2 - \sum_{1 \leq h \neq r \leq i} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&- \sum_{h=1}^{h=i} \sum_{r=i+1}^{r=j} (t_h - t_{h-1})(t_r - t_{r-1}) \\
&= t_i - t_i t_j = \min(t_i, t_j) - t_i t_j.
\end{aligned}$$

Ceci achève la preuve.

5. Principe d'invariance

Soit X_1, X, \dots une suite iid de variables aléatoires centrées avec $E|X_i|^2 < \infty$ et soit pour $n \geq 1$,

$$S_n = X_1 + \dots + X_n, n \geq 1.$$

Posons pour $0 \leq t \leq 1$,

$$S_n(t) = \frac{S_{[nt]}}{\sqrt{n}}$$

Nous allons étudier la loi limite du processus $\{S_n(t), 0 \leq t \leq 1\}$ en distribution finie.

Pour cela, soit $0 = t_0 < t_2 < \dots < t_k = 1$, $k \geq 1$. Nous avons

PROPOSITION 4. *The sequence of finite distributions*

$$\left(\frac{S_{[nt_j]}}{\sqrt{n}}, 1 \leq j \leq k \right), \quad n \geq 1,$$

weakly converges to k -dimensional centered Gaussian vector of variance-covariance matrix

$$(\min(t_i, t_j))_{1 \leq i, j \leq k}.$$

Proof. Nous avons

$$\left\{ \begin{array}{l} Y_n(t_1) = X_n(t_1) - X_n(t_0) = \frac{1}{\sqrt{n}} \sum_{[nt_0] < j \leq [nt_1]} X_j \\ \dots \\ Y_n(t_i) = X_n(t_i) - X_n(t_{i-1}) = \frac{1}{\sqrt{n}} \sum_{[nt_{i-1}] < j \leq [nt_i]} X_j \\ \dots \\ Y_n(t_k) = X_n(t_k) - X_n(t_{k-1}) = \frac{1}{\sqrt{n}} \sum_{[nt_{k-1}] < j \leq [nt_k]} X_j \end{array} \right\}.$$

Nous constatons que les variables $Y_n(t_i)$ sont indépendantes et pour chaque t_i , nous pouvons appliquer le théorème central limite dans \mathbb{R} ,

$$Y_n(t_i) = \frac{1}{\sqrt{n}} \sum_{[nt_{i-1}] < j \leq [nt_i]} X_j \rightarrow N(0, t_i - t_{i-1})$$

Dès lors, pour tout $u = (u_1, \dots, u_k) \in \mathbb{R}^k$

$$\mathbb{E}(\exp(\sum_{1 \leq i \leq k} Y_n(t_i) u_i)) = \prod_{1 \leq i \leq k} \mathbb{E}(\exp(Y_n(t_i) u_i)) \rightarrow \prod_{1 \leq i \leq k} e^{\frac{1}{2} u_i^2 / (t_i - t_{i-1})}.$$

Donc le vecteur $Y_n = (Y_n(t_i), 1 \leq i \leq k)$ tend vers un vecteur gaussien Z à composantes indépendantes et dont le $i^{\text{ème}}$ composante a la variance $t_i - t_{i-1}$.

Mais le vecteur $X_n = (X_n(t_i), 1 \leq i \leq k)$ est une transformation linéaire de Y_n de la forme

$$X_n = AY_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ 1 & \dots & 1 & 0 \\ 1 & \dots & \dots & 1 \end{pmatrix} Y_n$$

avec

$$A_{ij} = 1_{(i \leq j)}.$$

Alors X_n converge en loi vers le vecteur gaussien $V = AZ$, de sorte que

$$V_i = Z_1 + \dots + Z_i$$

et

$$Z_i = V_i - V_{i-1}.$$

Donc pour tout $1 \leq i \leq k$,

$$\mathbb{E}(V_i^2) = \sum_{1 \leq j \leq i} \mathbb{E}(Z_j^2) = \sum_{1 \leq j \leq i} (t_j - t_{j-1}) = t_i.$$

Et pour tout $1 \leq i \leq j \leq k$,

$$\mathbb{E}(V_i V_j) = \mathbb{E}(V_i(V_i + (V_j - V_i)))$$

et

$$= \mathbb{E}(V_i^2) + \mathbb{E}(V_i(V_j - V_i)).$$

Puisque

$$V_i = Z_1 + \dots + Z_i$$

et

$$Z_i = V_i - V_{i-1}$$

sont indépendants et centrés, il vient que

$$\mathbb{E}(V_i V_j) = \mathbb{E}(V_i) = t_i = t_i \wedge t_j.$$

Cela suffit pour démontrer le résultat.

CHAPTER 2

Théorie de la Convergence Vague

1. Introduction

Dans ce chapitre, nous traitons de la théorie unifiée de la convergence vague par sa caractérisation fonctionnelle. Nous souhaitons nous limiter dans ce texte à l'étude de cette théorie sur des mesures de probabilité sur \mathbb{R}^k , $k \geq 1$.

Cependant, il arrive que les preuves soient exactement les mêmes que le cas général pour des mesures de probabilité sur un espace métrique (S, d) muni de sa σ -algèbre borélienne $\mathcal{B}(S)$, ne faisant intervenir que les propriétés générales de la métrique. Dans de telles situations, nous énonçons la propriété et les preuves dans le cas général.

Pour traiter de la tension de suites de mesures de probabilité, c'est-à-dire, de l'existence pour ces suites, de sous-suites convergentes au sens vague, le traitement sera fait spécifiquement à \mathbb{R}^k par le biais du théorème de Helly-Bray.

Comme dans toute théorie limite, il faut nécessairement traiter de l'unicité de la limite, de critères de convergence, de compacité relative et de transformations éventuelles. Pour compacité relative, nous parlerons plutôt de tension.

2. Définition, Unicité et Théorème Portmanteau

DEFINITION 1. *La suite d'applications mesurables $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, \mathcal{B}(S))$ converge vaguement vers $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (S, \mathcal{B}(S))$ mesurable ssi pour toute fonction $f : S \mapsto \mathbb{R}$, continue et bornée (noté $f \in \mathcal{C}_b(S)$),*

$$(2.1) \quad \mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X) \text{ quand } n \rightarrow +\infty.$$

Nous remarquons que les espaces de départ n'ont aucune importance dans cette théorie, d'où le nom de convergence vague. Notons $L = \mathbb{P}_X = \mathbb{P}_\infty \circ X^{-1}$, la loi de X définie

$$\forall B \in \mathcal{B}_\infty(S), L(B) = \mathbb{P}_\infty(X^{-1}(B)) = \mathbb{P}(X \in B),$$

et pour tout $n \geq 1$, $\mathbb{P}^{(n)}$ la loi de probabilité de X_n définie par

$$\forall B \in \mathcal{B}(S), \mathbb{P}^{(n)}(B) = \mathbb{P}_n(X_n^{-1}(B)) = \mathbb{P}_n(X_n \in B).$$

La définition dit que X_n converge vaguement vers X si et seulement si pour tout $f \in \mathcal{C}_b(S)$,

$$\int_S f(x) d\mathbb{P}^{(n)}(x) \rightarrow \int_S f(x) dL(x) \text{ quand } n \rightarrow +\infty.$$

On pourrait ainsi remplacer (2.1) par

$$(2.2) \quad \mathbb{E}f(X_n) \rightarrow \int_S f dL \text{ quand } n \rightarrow +\infty,$$

et dire que la suite $(X_n)_{n \geq 1}$ converge vaguement vers une probabilité L . Dans la suite, nous utiliserons les deux terminologies.

Faites attention à ce point important. Il est aussi important de voir que les symboles d'espérance mathématique dans (2.1) dépendent des mesures de probabilité qu'ils utilisent. En conséquence, ils doivent être labellisés par exemple sous la forme,

$$\mathbb{E}_\infty(f(X)) = \int f(X) d\mathbb{P}_\infty, \quad \mathbb{E}_n(f(X_n)) = \int f(X_n) d\mathbb{P}_n, \quad n \geq 1.$$

avec $\mathbb{E}_\infty = \mathbb{E}_{\mathbb{P}_\infty}$ et $\mathbb{E}_n = \mathbb{E}_{\mathbb{P}_n}$. Mais, nous avons choisi d'omettre ces labels en vue de garder les notations simples. Nous les utiliserons lorsque cela est nécessaire. Ainsi, on ne doit utiliser la linéarité de l'espérance ou effectuer des opérations sur les espérances que lorsque les espaces de probabilité de départ sont les mêmes, comme dans la section 6.

Nous allons montrer que la limite est unique en distribution dans le sens suivant.

PROPOSITION 5. *Soit une suite d'applications mesurables $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, \mathcal{B}(S))$, et soit \mathbb{Q}_1 et \mathbb{Q}_2 deux probabilités sur $(S, \mathcal{B}(S))$. Supposons que X_n converge vaguement à la fois vers \mathbb{Q}_1 et \mathbb{Q}_2 . Alors, nécessairement, nous avons*

$$\mathbb{Q}_1 = \mathbb{Q}_2.$$

PREUVE. Supposons que X_n converge vaguement à la fois vers \mathbb{Q}_1 et \mathbb{Q}_2 . Pour montrer que $\mathbb{P}_1 = \mathbb{P}_2$, il suffit de montrer qu'elles ont égales sur la classe des ouverts Θ de (S, d) . En effet Θ est un π -system qui engendre $\mathcal{B}(S)$. Alors deux probabilités sur $(S, \mathcal{B}(S))$ qui sont égales sur Θ , sont par ailleurs égales sur $\mathcal{B}(S)$.

Soit G un ouvert de S . Pour tout entier $m \geq 1$, posons $f_m(x) = \min(m d(x, G^c), 1)$. Pour tout m , la fonction f_m est à valeurs dans $[0, 1]$, donc bornée. Puisque G^c est fermée, on a

$$d(x, G^c) = \begin{cases} > 0 & \text{si } x \in G \\ 0 & \text{si } x \in G^c \end{cases}.$$

Montrons que f_m est lipschitzienne. Evaluons $|f_m(x) - f_m(y)|$ selon trois cas.

Cas 1. $(x, y) \in (G^c)^2$. Donc

$$|f_m(x) - f_m(y)| = 0 \leq m d(x, y).$$

Cas 2. $x \in G$ et $y \in G^c$ (ou en permutant les rôles de x et y). On a

$$|f_m(x) - f_m(y)| = |\min(m d(x, G^c), 1)| \leq m d(x, G^c) \leq m d(x, y),$$

par définition même de $d(x, G^c) = \inf\{d(x, z), z \in G^c\}$.

Cas 3. Soit $(x, y) \in G^2$. On a, en utilisant la propriété (7.5) de la section appendice 7

$$\begin{aligned} |f_m(x) - f_m(y)| &= |\min(m d(x, G^c), 1) - \min(m d(y, G^c), 1)| \leq |m d(x, G^c) - m d(y, G^c)|, \\ &\leq m d(x, y) \end{aligned}$$

par l'inégalité triangulaire bis. Donc f_m est m -lipschitzienne. De plus

$$f_m \uparrow 1_G \text{ quand } m \uparrow \infty.$$

En effet, si $x \in G^c$, $f_m(x) = 0 \uparrow 0 = 1_G(x)$. Si $x \in G$, $d(x, G^c) > 0$ et $m d(x, G^c) \uparrow \infty$. Pour m assez grand,

$$(2.3) \quad f_m(x) = 1 \uparrow 1_G(x)$$

En résumé, chaque fonction f_m est lipschitzienne, bornée, positive.

Une fois que nous avons les propriétés de f_m , revenons à notre hypothèse à savoir que X_n converge vaguement à la fois vers \mathbb{P}_1 et \mathbb{P}_2 ,

c'est-à-dire que pour toute fonction $f : S \mapsto \mathbb{R}$, continue et bornée, nous avons, quand $n \rightarrow +\infty$,

$$(2.4) \quad \mathbb{E}f(X_n) \rightarrow \int f d\mathbb{Q}_1 \quad \text{and} \quad \mathbb{E}f(X_n) \rightarrow \int f d\mathbb{Q}_2.$$

Par l'unicité de la limite de suites réelles, nous avons

$$\forall (f \in C_b(S)), \int f d\mathbb{Q}_1 = \int f d\mathbb{Q}_2.$$

On peut donc appliquer cette égalité pour toutes les fonctions f_m définies ci-haut. Cela nous donnera

$$\forall (m \geq 1), \int f_m d\mathbb{Q}_1 = \int f_m d\mathbb{Q}_2.$$

En faisant croître m vers l'infini, en utilisant (2.3) et en appliquant le Théorème de convergence monotone, nous obtenons

$$\int 1_G d\mathbb{Q}_1 = \int 1_G d\mathbb{Q}_2,$$

signifiant

$$\mathbb{Q}_1(G) = \mathbb{Q}_2(G).$$

Puisque G a été arbitrairement fixé, il s'ensuit que l'égalité est vraie pour tous les ouverts de S . Alors, par nos remarques liminaires, nous avons bien $\mathbb{Q}_1 = \mathbb{Q}_2$.

Notation. Lorsque $(X_n)_{n \geq 1}$ converge vaguement X , nous utilisons la notation principale

$$X_n \rightsquigarrow X \text{ quand } n \rightarrow +\infty,$$

mais il nous arrivera aussi d'utiliser ces deux autres : $X_n \rightarrow_{\mathcal{L}} X$ (pour convergence en loi) or $X_n \rightarrow_d X$ (pour convergence en distribution) ou $X_n \rightarrow_w X$ (pour convergence vague, *weakly* en Anglais).

Nous avons maintenant besoin de la caractérisation de cette convergence. Selon les besoins, on peut avoir besoin d'autres angles d'attaque, pour l'établir.

THÉORÈME 2. (*Portmanteau*). *La suite d'applications $X_n : (\Omega_n, \mathcal{A}_n, P_n) \mapsto (S, B(S))$ converge vaguement vers une probabilité L ssi*

(ii) Pour tout ouvert G de S ,

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq L(G).$$

(iii) Pour tout fermé F de S ,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq L(F).$$

(iv) Pour toute fonction f semi-continue inférieurement (s.c.i) et minorée,

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \int f \, dL.$$

(v) Pour toute fonction f semi-continue supérieurement (s.c.s) et majorée,

$$\limsup_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \int f \, dL.$$

(vi) Pour tout borélien B de S tel que $L(\partial B) = 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = L(B).$$

(vii) Pour toute fonction f positive, bornée et lipschitzienne.

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \int f \, dL.$$

Avant de commencer la preuve, rappelons que ∂B est la frontière de B . Si $L(\partial B) = 0$, on dit que B est L -continue. Quant aux fonctions semi-continues, nous ferons un rappel dans la sous-section 7.2 de la section annexe 7.

PREUVE.

Les points (ii) et (iii) sont équivalents par complémentation. De même pour les points (iv) et (v) en passant de f à $-f$ et en utilisant les propriétés vues dans la sous-section 7.2 de la section annexe 7 ci-bas. Maintenant notons (i) la formule (2.1) de la définition de la convergence vague. L'implication (i) \Rightarrow (vii) est évidente car une fonction lipschitzienne (de paramètre k), c'est à dire, telle que

$$\forall (x, y) \in S^2, |f(x) - f(y)| \leq k \, d(x, y),$$

est continue.

Prouvons (vii) \Rightarrow (ii). Soit G un ouvert de S . Pour tout entier $m \geq 1$, posons $f_m(x) = \min(m d(x, G^c), 1)$. Pour tout m , la fonction f_m est à valeurs dans $[0,1]$, donc bornée. Nous avons déjà rencontré cette fonction dans la preuve de la proposition 5. Nous savons que chaque m , f_m est lipschitzienne, bornée, positive et

$$f_m \uparrow 1_G \text{ quand } m \uparrow \infty.$$

Nous avons pour tout $n \geq 1$ et pour tout $m \geq 1$,

$$\mathbb{E}(1_G(X_n)) \geq \mathbb{E}f_m(X_n).$$

Appliquons (vii) pour avoir

$$(2.5) \quad \liminf_{n \rightarrow +\infty} \mathbb{E}(1_G(X_n)) \geq \liminf_{n \rightarrow +\infty} \mathbb{E}f_m(X_n) \geq \int f_m dL.$$

Or pour toute mesure de probabilité et pour toute partie mesurable

$$\mathbb{E}_{\mathbb{Q}}(1_B) = \mathbb{Q}(B)$$

Pour $B = 1_{X_n^{-1}(G)} = 1_{(X_n \in G)}$, et en passant à la limite sur m et en utilisant le théorème de convergence monotone à (2.5), on obtient

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq \int 1_G dL = L(G)$$

Donc (ii) est démontré.

Prouvons que : (ii) \Rightarrow (iv).

Soit (ii) vraie. Soit f une fonction semi-continue inférieurement minorée par M . Nous pouvons prouver (iv) pour $f - M = g$ positive, qui est encore s.c.i. Alors les ensembles $(g \leq c)$ sont fermés selon la proposition 19 de la section annexe 7. Posons pour $m \geq 1$ fixé.

$$G_i = \{g > i/m\}, \quad i \geq 1.$$

et

$$g_m = \frac{1}{m} \sum_{i=1}^{m^2} 1_{G_i}.$$

Les ensembles G_i sont ouverts car g est s.c.i. Remarquons que

$$(2.6) \quad g_m(x) = \frac{i}{m} \text{ pour } \frac{i}{m} < g(x) \leq \frac{i+1}{m}, \text{ pour } i = 1, \dots, m^2 - 1$$

et

$$g_m(x) = m \text{ pour } g(x) > m.$$

Donc, nous avons

$$g_m \leq g.$$

De plus, d'après (2.6), il est vrai que

$$|g_m(x) - g(m)| \leq 1/m \text{ pour } g(x) \leq m.$$

On a

$$g(X_n) \geq g_m(X_n) = \frac{1}{m} \sum_{i=1}^{m^2} 1_{G_i}(X_n) = \frac{1}{m} \sum_{i=1}^{m^2} 1_{(X_n \in G_i)},$$

signifiant

$$(2.7) \quad \mathbb{E}g(X_n) \geq \mathbb{E}g_m(X_n) = \frac{1}{m} \mathbb{E} \cdot \sum_{i=1}^{m^2} 1_{(X_n \in G_i)}.$$

Donc (2.7) donne

$$\mathbb{E}g(X_n) \geq \mathbb{E}g_m(X_n) \geq \frac{1}{m} \sum_{i=1}^{m^2} \mathbb{E}1_{(X_n \in G_i)} = \frac{1}{m} \sum_{i=1}^{m^2} \mathbb{P}_n(X_n \in G_i).$$

En passant à la limite sur n , et en appliquant (ii), on aura

$$\liminf_{n \rightarrow +\infty} \mathbb{E}g(X_n) \geq \liminf_{n \rightarrow +\infty} \mathbb{E}g_m(X_n) \geq \frac{1}{m} \sum_{i=1}^{m^2} L(G_i) = \int g_m \, dL \geq \int_{(g \leq m)} g_m \, dL,$$

et donc

$$\geq \int_{(g \leq m)} g \, dL + \int_{(g \leq m)} (g_m - g) \, dL.$$

Quand $m \rightarrow \infty$

$$\int_{(g \leq m)} g \, dL \rightarrow \int g \, dL$$

et

$$|\int_{(g \leq m)} (g_m - g) dL| \leq L(S)/m \rightarrow 0.$$

D'où

$$\liminf_{n \rightarrow +\infty} \mathbb{E}g(X_n) \geq \int g dL.$$

Maintenant, en remplaçant g par $f - M$, la même formule reste vraie, par simplification de M . Donc (iv) vraie.

Prouvons (ii) \Rightarrow (vi). Rappelons que $\partial B = \overline{B} - \text{int}(B)$, autrement dit, B est la différence entre la fermeture de B et de son intérieur. Donc, puisque $\text{int}(B) \subseteq B \subseteq \overline{B}$

$$(2.8) \quad L(\partial B) = L(\text{int}(B)) - L(\overline{B}) = 0 \Rightarrow L(\text{int}(B)) = L(\overline{B}) = L(B)$$

Puisque $\text{int}(B)$ est ouvert et que \overline{B} est fermé, on peut utiliser (ii) et (iii) à la fois pour avoir

$$(2.9) \quad L(\text{int}(B)) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \text{int}(B)) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B),$$

$$(2.10) \quad \leq \mathbb{P}_n(X_n \in \overline{B}) \leq L(\text{int}(B),$$

D'où, par (2.8)

$$L(B) = \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B).$$

Ce qui était à démontrer.

Prouvons que (vi) \Rightarrow (iii).

Soit (vi) vraie et soit F un fermé de S . Posons $(F_\varepsilon) = \{x, d(x, F) \leq \varepsilon\}$ pour $\varepsilon \geq 0$. On

$$F \subseteq F(\varepsilon)$$

et

$$F(\varepsilon) \downarrow F \text{ pour } \varepsilon \downarrow 0$$

Maintenant $\partial F(\varepsilon) \subseteq \{x, d(x, F) = \varepsilon\}$. Donc les ensembles $\partial F(\varepsilon)$ sont disjoints, donc au plus un nombre dénombrable d'ensembles parmi eux,

ont une probabilité non nulle (voir Proposition 7.3 de la section annexe 7 ci-bas). Donc on peut trouver une suite $\varepsilon_n \downarrow 0$ telle que pour tout n ,

$$L(\partial F(\varepsilon_n)) = 0.$$

Pour n fixé, $F \subseteq F(\varepsilon_n)$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F(\varepsilon_n))$$

et par application de (vi)

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F(\varepsilon_n)) = L(F(\varepsilon_n)).$$

Maintenant, en passant à la limite quand $n \uparrow \infty$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in F) \leq L(F),$$

ce qui est bien (iii).

Prouvons $(iv) \Rightarrow (i)$

Si (iv) est vraie, alors (v) est vraie. Donc une fonction f continue et bornée, est à la fois s.c.i. et minorée, et s.c.s. et majorée, on aura

$$\int f dL \leq \liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \limsup_{n \rightarrow +\infty} \sup \mathbb{E}f(X_n) \leq \int f dL$$

D'où

$$\int f dL = \lim_{n \rightarrow +\infty} \mathbb{E}f(X_n) = \lim_{n \rightarrow +\infty} \mathbb{E}f(X_n).$$

En résumé nous avons montré le théorème par ce schéma

$$\begin{array}{ccccccc} (i) & \Rightarrow & (vii) & \Rightarrow & (ii) & \Leftrightarrow & (iii) \\ \uparrow & & & & \downarrow & & \uparrow \\ (v) & \Leftrightarrow & (iv) & = & (iv) & (vi) & = & (vi) \end{array}$$

qui montre que les six assertions sont équivalentes entre elles.

3. Transformations continues

Soit une suite $(X_n)_{n \geq 1}$ à valeurs dans l'espace métrique (S, d) qui converge vaguement vers X et soit une application g de S dans un autre espace métrique (D, r) . Alors, la suite $g(X_n)_{n \geq 1}$ converge-t-elle vaguement vers $g(X)$?

Pour répondre partiellement à cette question, supposons que g soit continue. Alors, il est évident que si $f \in C_b(D)$, alors $f \circ g \in C_b(S)$.

Donc, nous avons

$$\forall f \in C_b(S), \mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$$

et puisque $f \in C_b(D) \Rightarrow f \circ g \in C_b(S)$, cela mène à :

$$\forall f \in C_b(D), \mathbb{E}f \circ g(X_n) \rightarrow \mathbb{E}f \circ g(X).$$

D'où la proposition :

PROPOSITION 6. *Soit une suite d'applications $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, B(S))$ convergeant vaguement vers la variable aléatoire $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (S, B(S))$ (ou vers la probabilité L sur S), et $g : S \mapsto D$ une application continue. Alors*

$$g(X_n) \rightarrow_w g(X).$$

ou par une autre écriture,

$$g(X_n) \rightarrow_w L \circ g^{-1}$$

Cette proposition est très importante. Mais, on a plus que cela. En effet, on n'a pas besoin de la continuité sur tout l'ensemble S . Il suffira que l'ensemble des points de discontinuité de g soit de mesure nulle par rapport à L . Soit $\text{discont}(g)$ l'ensemble mesurable des points de discontinuité de g . D'après le lemme 3 de la Section Appendice 7 ci-bas, cet ensemble est mesurable.

Nous avons le résultat général suivant.

PROPOSITION 7. *Soit une suite d'applications $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (S, B(S))$ convergeant vaguement vers la variable aléatoire $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (S, B(S))$ (ou vers la probabilité L sur S), et $g : S \mapsto D$ une application telle que $L(\text{discont}(g)) = \mathbb{P}(X \in \text{discont}(g)) = 0$. Alors*

$$g(X_n) \rightarrow_w g(X)$$

ou par une autre écriture,

$$g(X_n) \rightarrow_w L \circ g^{-1}.$$

PREUVE. Soit $X_n \rightarrow_w L$ avec $L(\text{discont}(g)) = 0$. Soit F une partie fermée de D . Montrons que nous avons

$$(3.1) \quad \overline{g^{-1}(F)} \subseteq g^{-1}(F) \cup \text{discont}(g).$$

En effet soit $x \in \overline{g^{-1}(F)}$. Donc il existe une suite $(y_n)_{n \geq 1} \in g^{-1}(F)$ telle que $y_n \rightarrow x$ et pour tout $n \geq 1$, $g(y_n) \in F$. Alors *ou bien*

$$x \in \text{discont}(g),$$

ou bien x est un point de continuité. Dans ce dernier cas, $(y_n \rightarrow x, g(y_n) \in F) \Rightarrow g(y_n) \rightarrow g(x) \in F$ puisque F est fermé et donc :

$$x \in g^{-1}(F)$$

Nous voyons donc que (3.1) est établie. Combinons cette formule avec le point (iii) du théorème Portmanteau 2. Nous avons

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(g(X_n) \in F) = \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in g^{-1}(F)) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \overline{g^{-1}(F)})$$

et

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \overline{g^{-1}(F)}) \leq L(\overline{g^{-1}(F)}) \leq L(g^{-1}(F)) + L(\text{discont}(g)).$$

Ce qui donne

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(g(X_n) \in F) \leq L \circ g^{-1}(F).$$

D'où le résultat cherché.

4. Cas particulier de \mathbb{R}^k , $k \geq 1$

Intéressons nous au cas particulier $S = \mathbb{R}^k$. Avant d'aller plus loin, rappelons les caractérisations de lois de probabilités sur $S = \mathbb{R}^k$.

Pour commencer, faisons cette précision.

Terminologie Dans toute la partie traitant de \mathbb{R}^k , un vecteur aléatoire de dimension $k \geq 1$ est simplement une application mesurable définie

sur un espace mesurable donné à valeurs dans \mathbb{R}^k .

Ensuite, adoptons ces notations relatives aux coordonnées d'un vecteur X et d'une suite de vecteurs $(X_n)_{n \geq 1}$:

$$X = \begin{bmatrix} X_1 \\ \cdots \\ X_k \end{bmatrix},$$

$$X_n = \begin{bmatrix} X_1^{(n)} \\ \cdots \\ X_k^{(n)} \end{bmatrix}.$$

Nous arpellons (voir [8] dans notre série) que la loi de probabilité d'un vecteur aléatoire X de \mathbb{R}^k est caractérisée par sa *fonction de répartition*, définie par

$${}^t(t_1, t_2, \dots, t_k) \mapsto F_X(t_1, t_2, \dots, t_k) = \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_k \leq t_k).$$

Elle est aussi caractérisée par sa *fonction caractéristique* donnée par :

$${}^t(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k) = \mathbb{E}(\exp(\sum_j^k i u_j X_j))$$

Si sa *fonction des moments* donnée

$${}^t(u_1, u_2, \dots, u_k) \mapsto \Psi_X(u_1, u_2, \dots, u_k) = \mathbb{E}(\exp(\sum_j^k u_j X_j)),$$

existe autour d'un voisinage du vecteur nul, alors elle définit la loi de probabilité de X de manière unique.

Il en est de même pour *densité de probabilité* par rapport à la mesure de Lebesgues sur \mathbb{R}^k . Celle-ci existe si et seulement si

$${}^t(t_1, t_2, \dots, t_k) \mapsto f_X(t_1, t_2, \dots, t_k) = \frac{\partial^{(k)} F_X(t_1, t_2, \dots, t_k)}{\partial t_1 \partial t_2 \cdots \partial t_k},$$

presque partout - en (t_1, t_2, \dots, t_k) - par rapport à la mesure de Lebesgues sur \mathbb{R}^k .

Il est remarquable que ces mêmes déterminants jouent aussi les grands rôles en convergence vague

Avec les notations déjà introduites, nous avons les propositions suivantes.

PROPOSITION 8. *Une suite de variables aléatoires $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ converge vaguement vers la variable aléatoire $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, alors pour tout point $t = (t_1, t_2, \dots, t_k)$ de continuité de F_X ou de L ,*

$$(4.1) \quad \mathbb{P}_n(X_n \in \prod_{i=1}^k]-\infty, t_i]) \rightarrow F_X(t_1, t_2, \dots, t_k).$$

PREUVE. Soit la fonction de répartition de X

$$\begin{aligned} F_X(t_1, t_2, \dots, t_k) &= \mathbb{P}_\infty(X_1 \leq t_1, X_2 \leq t_2, \dots, X_k \leq t_k) \\ &= \mathbb{P}_\infty(X \in \prod_{i=1}^k]-\infty, t_i]) \end{aligned}$$

Notons $t = (t_1, \dots, t_k)$ et $t(n) = (t_1(n), t_2(n), \dots, t_k(n))$. On dira que $t(n) \uparrow t$ (resp $t(n) \downarrow t$) ssi

$$\forall (1 \leq i \leq k), t_i(n) \uparrow t_i \text{ (resp. } t_i(n) \downarrow t_i)$$

Posons $A(t) = \prod_{i=1}^k]-\infty, t_i]$. Nous avons quand $n \uparrow \infty$,

$$A(t(n)) \downarrow A(t)$$

et donc, par la limite monotone des probabilités,

$$F_X(t) = \mathbb{P}_\infty(X \in A(t(n)) \downarrow \mathbb{P}_\infty(X \in A(t)) = F_X(t)$$

Par suite, F_X est continue à droite en tout t . Mais

$$A(t(n)) \uparrow A^+(t) = \prod_{i=1}^k]-\infty, t_i[$$

et par suite

$$F_X(t) = \mathbb{P}_\infty(X \in A(t(n)) \uparrow \mathbb{P}_\infty(X \in A^+(t))$$

Mais nous avons

$$\begin{aligned} (4.2) \quad D(t) &= A(t) \setminus A^+(t) \\ &= \{x = (x_1, \dots, x_k) \in A(t), \exists 1 \leq i \leq k, x_i = t_i\} \end{aligned}$$

Pour mieux comprendre cette formule, regardez la pour $k=1$

$$D(a) =] - \infty, a] \setminus] - \infty, a[= \{a\}$$

et pour $k=2$

$$D(a, b) =] - \infty, a] \times] - \infty, b] \setminus] - \infty, a[\times] - \infty, b[$$

$$= \{(x, y) \in] - \infty, a] \times] - \infty, b], x = a \text{ ou } y = b\}$$

D'où, si

$$(4.3) \quad \mathbb{P}_\infty(X \in D(t)) = L(D(t)) = 0$$

alors, quand $n \rightarrow \infty$,

$$\begin{aligned} F_X(t) &= \mathbb{P}_\infty(X \in A(t(n))) \uparrow \mathbb{P}_\infty(X \in A^+(t)) \\ &= \mathbb{P}_\infty(X \in A(t)) - \mathbb{P}_\infty(X \in D(t)) = F_X(t) \end{aligned}$$

Donc (4.3) est la condition de continuité de F_X en t . Mais la frontière de $A(t)$ est exactement $D(t)$, i.e.,

$$(4.4) \quad \partial A(t) = D(t).$$

Car l'intérieur de $A(t)$ est sûrement $A^+(t)$. Donc d'après le point (vi) du théorème Portmanteau, nous avons la partie directe de la proposition : quand $n \rightarrow +\infty$,

$$(X_n \rightarrow_w X) \implies (\mathbb{P}_n(X_n \in] - \infty, t]) \rightarrow F_X(t) \text{ pour } F_X \text{ continue en } t.$$

Ce qui finit la preuve.

PROPOSITION 9. *Une suite d'applications $X_n : (\Omega_n, \mathcal{A}_n, P_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ mesurables converge vaguement vers la variable aléatoire $X : (\Omega_\infty, \mathcal{A}_\infty, P_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, ssi pour tout point $t = (t_1, t_2, \dots, t_k)$ de continuité de F_X ,*

$$(4.5) \quad F_{X_n}(t) \rightarrow F_X(t_1, t_2, \dots, t_k)$$

PREUVE. Supposons que pour tout point $t = (t_1, t_2, \dots, t_k)$ de continuité de F_X , $F_{X_n}(t) \rightarrow F_X(t_1, t_2, \dots, t_k)$. Pour montrer que X_n converge vaguement, nous allons montrer le point (ii) du Théorème Portmanteau, c'est-à-dire, pour tout ouvert G de \mathbb{R}^k ,

$$\liminf_{n \rightarrow +\infty} P(X_n \in G) \geq P(X \in G).$$

Soit un ouvert G de \mathbb{R}^k . D'après la proposition 18 de la section appendice 7 ci-bas, G peut s'écrire sous la forme

$$G = \bigcup_{j \geq 1}]a^j, b^j]$$

avec $]a^j, b^j]$ F_X -continu, c'est-à-dire que pour tout point c telle que

$$c_i = a_i^{(j)} \text{ ou } c_i = b_i^{(j)},$$

c est un point de continuité de F_X . Grâce à la continuité de la probabilité \mathbb{P}_X , on peut trouver pour tout $\eta > 0$, un rang m tel que

$$(4.6) \quad \mathbb{P}_X(G) - \eta \leq \mathbb{P}_X\left(\bigcup_{j=1}^m]a^j, b^j]\right)$$

Posons $A_j =]a^j, b^j]$ la formule de Poincaré donne

$$(4.7) \quad \begin{aligned} \mathbb{P}_X\left(\bigcup_{j=1}^m A_j\right) &= \sum \mathbb{P}_X(A_j) - \sum \mathbb{P}_X(A_i A_j) + \sum \mathbb{P}_X(A_i A_j A_k) \\ &+ \dots + (-1)^{n+1} \mathbb{P}_X(A_1 A_2 \dots A_n) \end{aligned}$$

et

$$(4.8) \quad \begin{aligned} \mathbb{P}_{X_n}\left(\bigcup_{j=1}^m A_j\right) &= \sum \mathbb{P}_{X_n}(A_j) - \sum \mathbb{P}_{X_n}(A_i A_j) + \sum \mathbb{P}_{X_n}(A_i A_j A_k) \\ &+ \dots + (-1)^{n+1} \mathbb{P}_{X_n}(A_1 A_2 \dots A_n) \end{aligned}$$

Maintenant traitons chacun des termes de ces expressions. Prenons un terme quelconque

$$\mathbb{P}_X(A_{i_1} A_{i_2} \dots A_{i_k}).$$

Puisque la classe \mathcal{U} des intervalles F_X -continus est stable par intersection finie (Sous-section 7.1.2 de la section 7), cette intersection $A_{i_1} A_{i_2} \dots A_{i_k}$ est de type $]a, b] \in \mathcal{U}$ et la formule de la mesure de Lebesgue-Stieljes donne

$$\mathbb{P}_X(A_{i_1} A_{i_2} \dots A_{i_k}) = \sum_{\varepsilon \in \{0,1\}^k} (-1)^{\sum_{1 \leq i \leq k} \varepsilon_i} F_X(b + \varepsilon * (a - b))$$

et

$$\mathbb{P}_{X_n}(A_{i_1} A_{i_2} \dots A_{i_k}) = \sum_{\varepsilon \in \{0,1\}^k} (-1)^{\sum_{1 \leq i \leq k} \varepsilon_i} F_{X_n}(b + \varepsilon * (a - b)).$$

Ici les points $c = b + \varepsilon * (a - b)$ sont tous des points de continuité puisque $]a, b] \in \mathcal{U}$. Il suit de cela et de la convergence (4.5) que

$$\mathbb{P}_{X_n}(A_{i_1}A_{i_2}\dots A_{i_k}) \rightarrow \mathbb{P}_X(A_{i_1}A_{i_2}\dots A_{i_k}).$$

En opérant par terme dans (4.7) et (4.8), nous concluons que

$$\mathbb{P}_{X_n}\left(\bigcup_{j=1}^m A_j\right) \rightarrow \mathbb{P}_X\left(\bigcup_{j=1}^m A_j\right),$$

et par la suite,

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) &= \liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \bigcup_{j \geq 1}]a^j, b^j]) \\ &\geq \lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in \bigcup_{j=1}^m]a^j, b^j]) \geq \mathbb{P}_n(X \in G) - \eta \end{aligned}$$

pour tout $\eta > 0$. D'où

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq \mathbb{P}_\infty(X \in G).$$

pour tout G ouvert. Il s'en suit que

$$X_n \rightarrow_d X \text{ quand } n \rightarrow +\infty.$$

Passons aux fonctions caractéristiques. Nous avons :

PROPOSITION 10. *Une suite d'applications $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ mesurables converge vaguement vers la variable aléatoire $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, ssi pour tout point ${}^t(u_1, u_2, \dots, u_k) \in R^k$,*

$$\Phi_{X_n}(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k).$$

Remarque. La preuve que nous présentons ici, est basée sur le théorème de Stone-Weirstrass, qui est un élément important de la topologie des espaces de fonctions continues définies sur un compact. Ce théorème est rappelé dans la proposition 6 de la section 7, voir ci-bas. Cependant, une autre preuve beaucoup plus jolie à nos yeux, est donnée dans le théorème 10 du chapitre 3. Ce dernier est basé sur le concept de tension et le théorème de continuité de Lévy.

Proof. Rappelons la définition de la fonction caractéristique ainsi qu'il suit

$${}^t(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k) = \mathbb{E}(\exp(\sum_{j=1}^k i u_j X_j)),$$

qui peut être ré-écrite ainsi

$${}^t(u_1, u_2, \dots, u_k) \mapsto \exp\left(\sum_j^k i u_j X_j\right) = \cos\left(\sum_j^k i u_j X_j\right) + i \sin\left(\sum_j^k u_j X_j\right).$$

Cette fonction complexe a des composantes qui sont des fonctions bornées et continues de X et par définition,

$$\mathbb{E} \exp\left(\sum_j^k i u_j X_j\right) = \mathbb{E} \cos\left(\sum_j^k i u_j X_j\right) + i \mathbb{E} \sin\left(\sum_j^k u_j X_j\right).$$

Donc par la simple définition de la convergence vague, nous avons en tout ${}^t(u_1, u_2, \dots, u_k) \in \mathbb{R}^k$,

$$(4.9) \quad \Phi_{X_n}(u_1, u_2, \dots, u_k) \mapsto \Phi_X(u_1, u_2, \dots, u_k).$$

Ceci établit le sens direct de la preuve. Pour prouver le sens indirect one, nous en appelons au théorème de Stone-Weirstrass (Voir Proposition 6 de la sous-section 6 de la section 7, voir ci-bas).

A cet effet, posons $\Delta(a) = [-a, a]^k$, for $0 < a \in \mathbb{R}$. Notons par H la classe des fonctions continue bornées sur \mathbb{R}^k dont les éléments sont de la forme

$$(4.10) \quad {}^t(x_1, x_2, \dots, x_k) \mapsto \sum_{r=1}^m a_r \exp\left(\sum_j^k \pi n_{j,r} i x_j / a\right),$$

où les coefficients a_r sont des nombres complexes et les $n_{j,r}$ sont des entiers. En d'autres termes, les éléments de H sont des combinaisons linéaires de fonctions exponentielles de combinaisons linéaires de x_1, x_2, \dots, x_k . Remarquons que chaque fonction

$$x_j \mapsto \exp(\pi n_{j,r} i x_j / a)$$

est périodique de période $2a$, si bien que $h \in H$ attein ses valeurs dans $\Delta(a)$, et alors $\{h(x), x \in \mathbb{R}^k\} \subset \{h(x), x \in \Delta(a)\}$ et donc

$$\|h\| = \|h\|_{\Delta(a)}.$$

Pour tout $h \in H$, notons $h_{\Delta(b)}$ la restriction de h sur $\Delta(b)$, pour $b > 0$, et soit $H_b = \{h_{\Delta(b)}, h \in H\}$. Vérifions que pour tout $0 < b < a$, les conditions du théorème de Stone-Weirstrass sont vérifiées pour H_b .

(1) H contient les constantes. Soit d un nombre complexe quelconque. Dans (4.10), prenons $m = 1, a_1 = d$ et $n_{1,1} = n_{2,1} = \dots = n_{k,1} = 0$.

Nous obtenons $h(x) = a_1 = d$.

(2) H est stable par somme et produits finis. Cela est évident.

(3) Les conjugués \bar{h} des éléments h de H restent dans H .

(4) H sépare les points de \mathbb{R}^k . Pour le voir, prenons ${}^t(z_1, z_2, \dots, z_k) \neq {}^t(y_1, y_2, \dots, y_k)$ dans $\Delta(b)$. Donc au moins pour un indice $j_0 \in \{1, \dots, k\}$, nous avons $z_{j_0} \neq y_{j_0}$. Mais $|z_{j_0} - y_{j_0}| \leq 2b < 2a$. Définissons $h_0 \in H$ par

$$h_0(x) = e^{i\pi x_{j_0}/a}.$$

L'égalité $e^{i\pi z_{j_0}/a} = e^{i\pi y_{j_0}/a}$ aurait entraîné qu'il existe un entier $k \in \mathbb{Z}$ tel que $(z_{j_0} - y_{j_0}) = 2ka$, ce qui est impossible. Donc, nous avons

$$h_0(z) = e^{i\pi z_{j_0}/a} \neq e^{i\pi y_{j_0}/a} = h_0(y).$$

Ce qui implique que h_0 sépare ${}^t(z_1, z_2, \dots, z_k)$ et ${}^t(y_1, y_2, \dots, y_k)$.

Les conditions du théorème de Stone-Weirstrass sont vérifiées. Maintenant, supposons que (4.9) est vraie. Soit $f \in C_b(\mathbb{R}^k, \mathbb{R}) \subseteq C_b(\mathbb{R}^k, \mathbb{C})$. Soit f_b la restriction de f sur $\Delta(b)$ si bien que $f_b \in C(\Delta(b), \mathbb{C})$ où $b = b(a) < a$, est une fonction croissante non bornée de a ($b(a) = a/2$ for instance). A cet étape, nous appliquons le théorème de Stone-Weirstrass pour trouver, pour tout $\varepsilon > 0$, un élément h de H ,

$$h(x) = \sum_{r=1}^m a_r \exp\left(\sum_j^k i u_{r,j} x_j\right)$$

tel que

$$\sup_{x \in \Delta(b)} |f(x) - h(x)| = \|f - h\|_{\Delta(b)} \leq \varepsilon/3.$$

En appliquant (4.9), nous avons

$$\mathbb{E}(h(X_n)) \rightarrow \mathbb{E}(h(X)) \text{ as } n \rightarrow +\infty.$$

Soit donc un entier positif n_0 tel que pour tout $n \geq n_0$,

$$(4.11) \quad |\mathbb{E}(h(X_n)) - \mathbb{E}(h(X))| = \left| \int h \, d\mathbb{P}_n \circ X_n^{-1} - \int h \, d\mathbb{P} \circ X^{-1} \right| \leq \varepsilon/3$$

Nous avons

$$\begin{aligned} \mathbb{E}(f(X_n)) - \mathbb{E}(f(X)) &= \left(\int f \, d\mathbb{P}_n \circ X_n^{-1} - \int h \, d\mathbb{P}_n \circ X_n^{-1} \right) \\ &\quad + \left(\int h \, d\mathbb{P}_n \circ X_n^{-1} - \int_c h \, d\mathbb{P} \circ X^{-1} \right) \\ &\quad + \left(\int h \, d\mathbb{P} \circ X^{-1} - \int_c f \, d\mathbb{P} \circ X^{-1} \right). \end{aligned}$$

Le premier terme satisfait

$$\begin{aligned} \mathbb{E} \left| \int f \, d\mathbb{P}_n \circ X_n^{-1} - \int h \, d\mathbb{P}_n \circ X_n^{-1} \right| &\leq \int_{\Delta(b)} |f - h| \, d\mathbb{P}_n \circ X_n^{-1} \\ &\quad + \int_{\Delta^c(b)} |f - h| \, d\mathbb{P}_n \circ X_n^{-1} \\ (4.12) \qquad \qquad \qquad &\leq \varepsilon/3 + (\|f\| + \|h\|) \mathbb{P}_n(X_n \in \Delta^c(b)). \end{aligned}$$

Par la même méthode, le troisième terme aussi vérifie

$$(4.13) \quad \mathbb{E} \left| \int f \, d\mathbb{P}_\infty \circ X^{-1} - \int h \, d\mathbb{P}_\infty \circ X^{-1} \right| \leq \varepsilon/3 + (\|f\| + \|h\|) \mathbb{P}_\infty(X \in \Delta^c(b)).$$

En mettant ensemble les formules (4.11), (4.12) and (4.13), nous obtenons pour tout $n \geq n_0$

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq \varepsilon + (\|f\| + \|h\|)(\mathbb{P}_n(X_n \in \Delta^c(b)) + \mathbb{P}_\infty(X \in \Delta^c(b))).$$

Alors, pour tout $n \geq n_0$ fixé, nous faisons croître a vers $+\infty$. En conséquence $b(a) \uparrow +\infty$ et par la suite, $\mathbb{P}_n(X_n \in \Delta^c(b)) + \mathbb{P}_\infty(X \in \Delta^c(b)) \downarrow 0$. Alors pour tout $n \geq n_0$

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq \varepsilon$$

Ce qui prouve que

$$\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X)) \text{ as } n \rightarrow +\infty$$

et donc

$$X_n \rightsquigarrow X \text{ as } n \rightarrow +\infty.$$

En mettant ensemble les propositions (8), (9) et (10), nous obtenons le thorme complet Pormanteau dans \mathbb{R}^k .

THÉORÈME 3. *Let k be a positive integer. The sequence of random vectors $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $n \geq 1$, weakly converges to the random vector $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ if and only if one of these assertions holds.*

(i) *For any real-valued continuous and bounded function f defined on \mathbb{R}^k ,*

$$\lim_{n \rightarrow +\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X).$$

(ii) *For any open set G in \mathbb{R}^k ,*

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \geq \mathbb{P}_\infty(X \in G).$$

(iii) *For any closed set G of S , we have*

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in G) \leq \mathbb{P}_\infty(X \in G).$$

(iv) *For any inferior semi-continuous and bounded below function f , we have*

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X).$$

(v) *For any superior semi-continuous and bounded above function f , we have*

$$\limsup_{n \rightarrow +\infty} \mathbb{E}f(X_n) \leq \mathbb{E}f(X).$$

(vi) *For any Borel set B of S that is \mathbb{P}_X -continuous, that is $\mathbb{P}_\infty(X \in \partial B) = 0$, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{P}_n(X_n \in B) = \mathbb{P}_X(B) = \mathbb{P}_\infty(X \in B).$$

(vii) *For any nonnegative and bounded Lipschitz function f , we have*

$$\liminf_{n \rightarrow +\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X).$$

(viii) *For any continuity point $t = (t_1, t_2, \dots, t_k)$ of F_X , we have,*

$$F_{X_n}(t) \rightarrow F_X(t) \text{ as } n \rightarrow +\infty.$$

where for each $n \geq 1$, F_{X_n} is the distribution function of X_n and F_X that of X .

(ix) For any point $u = (u_1, u_2, \dots, u_k) \in \mathbb{R}^k$,

$$\Phi_{X_n}(u) \mapsto \Phi_X(u) \text{ as } n \rightarrow +\infty.$$

where for each $n \geq 1$, Φ_{X_n} is the characteristic function of X_n and Φ_X is that of X

Crière de Wold. La suite de vecteurs alatoires $\{X_n, n \geq 1\} \subset \mathbb{R}^k$ converge vaguement vers $X \in \mathbb{R}^k$, quand $n \rightarrow +\infty$ si et seulement si pour $a \in \mathbb{R}^k$, la suite $\{< a, X_n >, n \geq 1\} \subset \mathbb{R}$ converge veguement vers $X \in \mathbb{R}$ quand $n \rightarrow +\infty$.

Preuve. La preuve est rapide. Elle utilise les notations antérieures. Supposons que X_n converge faiblement vers X dans \mathbb{R}^k quand $n \rightarrow +\infty$. En utilisant la convergence des fonctions caratristiques, nous avons pour tout $u \in \mathbb{R}^k$,

$$\mathbb{E}(\exp(i < X_n, u >)) \rightarrow \mathbb{E}(\exp(i < X, u >)) \text{ quand } n \rightarrow +\infty.$$

Il s'en suit que $a \in \mathbb{R}^k$ et pour tout $t \in \mathbb{R}$, nous avons

$$(4.14) \quad \mathbb{E}(\exp(it < X_n, a >)) \rightarrow \mathbb{E}(\exp(it < X, a >)) \text{ quand } n \rightarrow +\infty.$$

c-a-d que, en prenant $u = ta$ dans la formule prcdente, et en notant $Z_n = < X_n, a >$ et $Z = < X, a >$,

$$\mathbb{E}(\exp(itZ_n)) \rightarrow \mathbb{E}(\exp(itZ)) \text{ quand } n \rightarrow +\infty.$$

Cela signifie que $Z_n \rightsquigarrow Z$, qui est gal $< a, X_n >$, converges vaguement vers $< a, X >$.

Inversement, supposons que pour tout $a \in \mathbb{R}^k$, la suite $\{< a, X_n >, n \geq 1\} \subset \mathbb{R}$ converge vaguement vers $X \in \mathbb{R}$ as $n \rightarrow +\infty$. Alors, en prenant $t = 1$ dans (4.14), nous obtenons pour tout $a = u \in \mathbb{R}^k$,

$$\mathbb{E}(\exp(i < X, u >)) \rightarrow \mathbb{E}(\exp(i < X, u >)) \text{ quand } n \rightarrow +\infty.$$

ce qui signifie que $X_n \rightsquigarrow +\infty$ quand $n \rightarrow +\infty$.

5. Théorème de Scheffé

Dans la section précédente, nous avons lié la convergence vague et quelques caractéristiques de variables aléatoires dans \mathbb{R}^k telles que la fonction de répartition et la fonction caractéristique. On peut se demander ce qu'il en est par rapport aux densités de probabilités par rapport à la mesure de Lebesgues dans \mathbb{R}^k . Le théorème de Scheffé(1947)

répond à cette préoccupation dans le cadre général. Enonçons-le d'abord.

THÉORÈME 4. (*Théorème de Scheffé*) Soit λ une mesure sur un espace mesurable (E, \mathcal{B}) . Et soit $p, (p_n)_{n \geq 1}$ des densités de probabilités par rapport à λ , c'est-à-dire des applications numériques définies sur E , positives, mesurables telles que

$$(5.1) \quad \forall n \geq 1, \int p_n d\lambda = \int p d\lambda = 1.$$

Si

$$p_n \rightarrow p, \quad \lambda - pp$$

alors

$$(5.2) \quad \sup_{B \in \mathcal{B}} \left| \int_B p_n d\lambda - \int_B p d\lambda \right| = \frac{1}{2} \int |p_n - p| d\lambda \rightarrow 0$$

PREUVE. Supposons que $p_n \rightarrow p, \lambda - pp$. Posons $\Delta_n = p - p_n$. Alors, (5.1) implique

$$\int \Delta_n d\lambda = 0.$$

Donc, pour $B \in \mathcal{B}$,

$$\int_{B^c} \Delta_n d\lambda = \int \Delta_n d\lambda - \int_B \Delta_n d\lambda = - \int_B \Delta_n d\lambda.$$

D'où

$$(5.3) \quad 2 \left| \int_B \Delta_n d\lambda \right| = \left| \int_B \Delta_n d\lambda \right| + \left| \int_{B^c} \Delta_n d\lambda \right| \leq \int_B |\Delta_n| d\lambda + \int_{B^c} |\Delta_n| d\lambda \leq \int |\Delta_n| d\lambda,$$

c'est-à-dire

$$(5.4) \quad \left| \int_B \Delta_n d\lambda \right| \leq \frac{1}{2} \int |\Delta_n| d\lambda.$$

En prenant $B = (\Delta_n \geq 0)$ dans (5.3), nous avons

$$2 \left| \int_B \Delta_n d\lambda \right| = \left| \int_B \Delta_n^+ d\lambda \right| + \left| \int_{B^c} -\Delta_n^- d\lambda \right| = \int \Delta_n^+ d\lambda + \int \Delta_n^- d\lambda = \int |\Delta_n| d\lambda.$$

En mettant ensemble les deux dernières formules, nous avons

$$(5.5) \quad \sup_{B \in \mathcal{B}} \left| \int_B p_n d\lambda - \int_B p d\lambda \right| = \frac{1}{2} \int |p_n - p| d\lambda.$$

Maintenant,

$$0 \leq \Delta_n^+ = \max(0, p - p_n) \leq p.$$

De plus,

$$\int \Delta_n^+ d\lambda = \int_{(\Delta_n \geq 0)} \Delta_n d\lambda = \int \Delta_n d\lambda - \int_{(\Delta_n \leq 0)} \Delta_n d\lambda = \int_{(\Delta_n \leq 0)} -\Delta_n d\lambda = \int \Delta_n^- d\lambda,$$

de sorte que

$$(5.6) \quad \int |\Delta_n| d\lambda = 2 \int \Delta_n^+ d\lambda$$

Appliquons le théorème de convergence dominée de Lebesgues à $0 \leq \Delta_n^+ \leq |\Delta_n| \rightarrow 0$ λ -p.p., $0 \leq \Delta_n^+ \leq p$. Nous aurons

$$\int \Delta_n^+ d\lambda \rightarrow 0,$$

en vertu de (5.5),

$$\sup_{B \in \mathcal{B}} \left| \int_B p_n d\lambda - \int_B p d\lambda \right| = \frac{1}{2} \int |p_n - p| d\lambda = \int \Delta_n^+ d\lambda \rightarrow 0$$

Le théorème de Scheffé peut alors s'appliquer aux densités de probabilité dans \mathbb{R}^k en particulier. Nous aurons :

PROPOSITION 11. (A) Soit $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ une suite de vecteurs aléatoires et $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ un autre vecteur aléatoire, tous de lois absolument continues par rapport à la mesure de Lebesgues sur \mathbb{R}^k notée λ_k . Soit, pour tout $n \geq 1$, f_{X_n} la densité de probabilité de X_n , et f_X celle de X . Supposons que :

$$f_{X_n} \rightarrow f_X, \quad \lambda_k - p.s., \quad \text{as } n \rightarrow +\infty.$$

Alors X_n converge vaguement vers X quand $n \rightarrow +\infty$.

(B) Soit $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ une suite vecteurs aléatoires valeurs discrètes $X : (\Omega_\infty, \mathcal{A}_\infty, \mathbb{P}_\infty) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ un autre vecteur aléatoires discret. Notons pour chaque n , le support dénombrable de X_n par D_n , c'est-à-dire :

$$\mathbb{P}_n(X_n \in D_n) = 1 \text{ et pour chaque } x \in D_n, \mathbb{P}_n(X_n = x) \neq 0$$

et par D_∞ le support dénombrable de X . Soit $D = D_\infty \cap \bigcap_{n \geq 1} D_n$ et soit ν la mesure de comptage sur D . Alors les densités de probabilité des vecteurs X_n et de X par rapport à ν sont définies sur D par

$$f_{X_n}(x) = \mathbb{P}_n(X_n = x), \quad n \geq 1, \quad f_X(x) = \mathbb{P}_\infty(X = x), \quad x \in D.$$

De plus, si

$$(\forall x \in D), f_{X_n}(x) \rightarrow f_X(x).$$

Alors X_n converge vaguement vers X .

6. Convergence vague et convergence en probabilité sur le même espace de probabilité

Cette section met la convergence vague à sa place dans le schéma de l'étude de convergence de suites de variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeur dans un espace métrique (S, d) .

Nous avons déjà vu que la convergence vague n'exige pas la connaissance des espaces de départ. Lorsque les suites étudiées sont sur le même espace, nous avons alors des relations intéressantes avec les autres types de convergence.

A l'inverse, il existe des résultats regroupés sous le manteau de Skorohod-Wichura-Dudley qui permettent de donner une version de convergence vague en convergence presque-sûre dans un espace convenablement choisi. La démonstration ne sera pas abordée dans cet ouvrage. Elle sera donnée dans l'ouvrage avancé de convergence vague. Ici, Il sera exposé et illustré seulement pour le cas de l'espace \mathbb{R} dans le chapitre 4.

Commençons par les définitions.

6.1. Définitions. Dans tout ce document, à l'exception de la section sur le théorème de Skorohod-Wichura, les suites de variables aléatoires $(X_n)_{n \geq 0}$, $(Y_n)_{n \geq 0}$, etc., et les variables aléatoires X , Y , etc. sont définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et sont à valeurs dans un espace métrique (S, d) . Nous aurons aussi à utiliser des constantes c dans S .

(a) Convergence presque-sûre.

La suite $(X_n)_{n \geq 0}$ converge presque-sûrement vers X , noté $X_n \longrightarrow X$, *a.s.*, si et seulement l'ensemble sur lequel $(X_n)_{n \geq 0}$ ne converge pas vers X est négligeable, c'est-à-dire que

$$\mathbb{P}(\{\omega \in \Omega, X_n \not\rightarrow X\}) = \mathbb{P}(\{\omega \in \Omega, d(X_n, X) \not\rightarrow 0\}) = 0.$$

Nous caractérisons cet ensemble aisément par

$$(X_n \not\rightarrow X) = \bigcup_{k \geq 1} \bigcap_{n \geq 0} \bigcup_{p \geq n} (d(X_p, X) > k^{-1}).$$

Ceci mène à la définition suivante : $(X_n)_{n \geq 0}$ converge presque-sûrement vers X si et seulement si :

$$(6.1) \quad \forall k \geq 1, \quad \mathbb{P} \left(\bigcap_{n \geq 0} \bigcup_{p \geq n} (d(X_p, X) > k^{-1}) \right) = 0.$$

(b) Convergence en probabilité.

La suite $(X_n)_{n \geq 0}$ converge en probabilité vers X , noté $X_n \longrightarrow_{\mathbb{P}} X$, *a.s.* si et seulement si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0.$$

Nous allons faire une brève comparaison entre ces deux types de convergence que vous connaissez déjà :

PROPOSITION 12. *Si $X_n \longrightarrow X$, *a.s.*, alors $X_n \longrightarrow_{\mathbb{P}} X$.*

Preuve. La preuve est classique. Il suffit de supposer qu'il y a convergence presque-sûre, c'est-à-dire que (6.1) a lieu et de voir que pour tout $k \geq 1$,

$$(d(X_n, X) > k^{-1}) \subset \bigcup_{p \geq n} (d(X_p, X) > k^{-1}) =: B_{n,k}.$$

Mais la suite $B_{n,k}$ (en n) est décroissante vers

$$\bigcap_{n \geq 0} \bigcup_{p \geq n} (d(X_p, X) > k^{-1}) =: B_k.$$

Donc pour tout $n \geq 0$ et pour tout $k \geq 1$,

$$\mathbb{P}(d(X_n, X) > k^{-1}) \leq \mathbb{P}(B_{n,k})$$

et par suite, par la continuité de la probabilité,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > k^{-1}) \leq \lim_{n \rightarrow +\infty} \mathbb{P}(B_{n,k}) = \mathbb{P}(B_k) = 0,$$

en appliquant (6.1).

6.2. Convergence vague et convergence en probabilité. Avant d'énoncer les propriétés, nous pouvons enrichir le Théorème Portmanteau par ce point supplémentaire.

LEMME 1. *La suite $(X_n)_{n \geq 0}$ converge vaguement vers X si et seulement si :*

(viii) *Pour toute fonction $f : S \rightarrow \mathbb{R}$ bornée et Lipschitzienne,*

$$\mathbb{E}f(X_n) \rightarrow f(X).$$

Preuve. Nous nous mettons dans le cadre de la démonstration du Théorème Portmanteau. Ce point (viii) entraîne le point (i) puisque f est continue bornée. De plus, (vii) est un cas particulier de (viii), ce qui fait que (viii) \implies (vii). En retour, si (vii) a lieu, on peut considérer l'infimum A de la fonction f et son supremum B et appliquer le point (vii) à la fonction $f - A$ puis à $-f + B$, ce qui entraînera le point (viii). On aura ainsi (i) \iff (vii) \iff (viii). Ce qui termine la preuve.

Nous allons maintenant énoncer un ensemble de propriétés.

(a) La convergence en probabilité entraîne la convergence vague

PROPOSITION 13. *Si $X_n \rightarrow_{\mathbb{P}} X$ implique que $X_n \rightsquigarrow X$*

Preuve. Supposons que $X_n \rightarrow_{\mathbb{P}} X$. Montrons que $X_n \rightsquigarrow X$ en utilisant le point (viii) du Lemme 1 ci-dessus. Soit donc une fonction f lipschitzienne de rapport ℓ et bornée par M . Nous avons pour tout $n \geq 0$,

$$|f(X_n) - f(X)| \leq \ell d(X_n, X).$$

Nous avons pour tout $n \geq 0$ et pour $\varepsilon > 0$,

$$\begin{aligned} |\mathbb{E}f(X_n) - \mathbb{E}f(X)| &\leq \mathbb{E}|f(X_n) - f(X)| \\ &\leq \int_{(d(X_n, X) \leq \varepsilon)} |f(X_n) - f(X)| d\mathbb{P} \\ &\quad + \int_{(d(X_n, X) > \varepsilon)} |f(X_n) - f(X)| d\mathbb{P}. \end{aligned}$$

Mais, pour tout $n \geq 0$ et pour $\varepsilon > 0$,

$$\int_{(d(X_n, X) \leq \varepsilon)} \ell d(X_n, X) d\mathbb{P} \leq \ell \varepsilon.$$

De plus, pour tout $n \geq 0$ et pour $\varepsilon > 0$,

$$\int_{(d(X_n, X) > \varepsilon)} |f(X_n) - f(X)| d\mathbb{P} \leq \int_{(d(X_n, X) > \varepsilon)} 2M d\mathbb{P} \leq 2M \mathbb{P}(d(X_n, X) > \varepsilon).$$

D'où, pour tout $n \geq 0$ et pour $\varepsilon > 0$,

$$|\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq \ell \varepsilon + 2M \mathbb{P}(d(X_n, X) > \varepsilon).$$

D'où pour tout $\varepsilon > 0$,

$$\limsup_{n \rightarrow +\infty} |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq \ell \varepsilon.$$

Nous obtenons, lorsque $\varepsilon \downarrow 0$.

$$\mathbb{E}f(X_n) \longrightarrow \mathbb{E}f(X).$$

Ce qui finit la preuve.

(b) La convergence vague et la convergence en probabilité vers une constante sont équivalentes

PROPOSITION 14. $X_n \rightarrow_{\mathbb{P}} c$ si et seulement si $X_n \rightsquigarrow c$

Preuve. Le sens $(X_n \rightarrow_{\mathbb{P}} c) \Rightarrow (X_n \rightsquigarrow c)$. Prouvons maintenant que $(X_n \rightsquigarrow c) \Rightarrow (X_n \rightarrow_{\mathbb{P}} c)$. Pour cela, supposons que $(X_n \rightsquigarrow c)$ et soit $\varepsilon > 0$. Le point (ii) du Théorème Portmanteau donne

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) < \varepsilon) &= \liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in B(c, \varepsilon)) \leq \mathbb{P}(c \in B(c, \varepsilon)) \\ &= \mathbb{P}(d(c, c) > \varepsilon) \\ &= \mathbb{P}(\Omega) = 1. \end{aligned}$$

D'où

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) \geq \varepsilon) = 1 - \liminf_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) < \varepsilon) \leq 1 - 1 = 0.$$

D'où, pour tout $0 < \varepsilon$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) > \varepsilon) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, c) \geq \varepsilon) = 0.$$

(c) Deux suites équivalentes en probabilité converge vaguement vers la même limite, s'il y a lieu.

PROPOSITION 15. Si $X_n \rightsquigarrow X$ et $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$, alors $Y_n \rightsquigarrow X$

Preuve. Supposons que $X_n \rightsquigarrow X$ et $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$. Montrons que $Y_n \rightsquigarrow X$ en utilisant le point (viii) du Lemma 1 ci-dessus. Soit une fonction f lipschitzienne de rapport ℓ et bornée par M . Nous aurons, Nous avons pour tout $n \geq 0$ et pour $\varepsilon > 0$,

$$\begin{aligned} |\mathbb{E}f(Y_n) - \mathbb{E}f(X)| &\leq \mathbb{E}|f(Y_n) - f(X)| \\ &\leq \mathbb{E}|f(X_n) - f(X)| + \mathbb{E}|f(Y_n) - f(X_n)|. \end{aligned}$$

Alors, en appliquant (viii) et se fondant sur le fait que $X_n \rightsquigarrow X$, on a

$$\limsup_{n \rightarrow +\infty} |\mathbb{E}f(Y_n) - \mathbb{E}f(X)| \leq \limsup_{n \rightarrow +\infty} \mathbb{E}|f(Y_n) - f(X_n)|.$$

Avec la même méthode utilisée dans la preuve de la proposition 13, nous aurons

$$\begin{aligned} \mathbb{E}|f(Y_n) - f(X_n)| &\leq \int_{(d(Y_n, X_n) \leq \varepsilon)} |f(Y_n) - f(X_n)| d\mathbb{P} \\ &\quad + \int_{(d(Y_n, X_n) > \varepsilon)} |f(Y_n) - f(X_n)| d\mathbb{P} \\ &\leq \ell\varepsilon + 2M \mathbb{P}(d(Y_n, X_n) > \varepsilon), \end{aligned}$$

ce qui tend vers zéro quand $n \rightarrow \infty$ puis $\varepsilon \downarrow 0$. Et par suite

$$\limsup_{n \rightarrow +\infty} |\mathbb{E}f(Y_n) - \mathbb{E}f(X)| = 0.$$

(d) Théorème de Slutsky.

Nous avons cet important outil de convergence vague.

PROPOSITION 16. Si $X_n \rightsquigarrow X$ et $Y_n \rightsquigarrow c$, alors $(X_n, Y_n) \rightsquigarrow (X, c)$

Preuve. Soit $X_n \rightsquigarrow X$ et $Y_n \xrightarrow{\mathbb{P}} c$. On veut montrer que $(X_n, Y_n) \rightsquigarrow (X, c)$. Remarquons d'abord que $Y_n \xrightarrow{\mathbb{P}} c$ puisque $Y_n \rightsquigarrow c$. De plus, sur S^2 munie de la métrique euclidienne

$$d_e((x', y'), (x'', y'')) = \sqrt{d(x', x'')^2 + d(y', y'')^2},$$

nous avons

$$d_e((X_n, Y_n), (X_n, c)) = d(Y_n, c).$$

Il s'en suit que pour $\varepsilon > 0$

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(d_e((X_n, Y_n), (X_n, c)) > \varepsilon) = \limsup_{n \rightarrow +\infty} \mathbb{P}(d(Y_n, c) > \varepsilon) = 0,$$

puisque $Y_n \rightarrow_{\mathbb{P}} c$. Ainsi $d_e((X_n, Y_n), (X_n, c)) \rightarrow_{\mathbb{P}} 0$. D'après la proposition de la sous-section XX, il suffit de montrer la convergence vague de (X_n, c) pour avoir celle de (X_n, Y_n) .

Maintenant, montrer la convergence vague de (X_n, c) vers (X, c) , revient à montrer que pour fonction $g(\cdot, \cdot)$ réelle continue et bornée définie sur S^2 , on a $\mathbb{E}g(X_n, c) \rightarrow \mathbb{E}g(X, c)$. Mais ceci vient du fait que, c étant fixée, la fonction $f(x) = g(x, c)$ est continue et bornée et $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$ puisque $X_n \rightsquigarrow X$. Mais $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$ équivaut à $\mathbb{E}g(X_n, c) \rightarrow \mathbb{E}g(X, c)$. Ce qui finit la preuve.

(e) Les convergence par coordonnées vague et en probabilité ne sont pas équivalentes.

PROPOSITION 17. Si $X_n \rightarrow_{\mathbb{P}} X$ et $Y_n \rightarrow_{\mathbb{P}} Y$, alors $(X_n, Y_n) \rightarrow_{\mathbb{P}} (X, Y)$

Preuve. Utilisons la distance de Manhattan sur S^2 :

$$d_m((x', y')x', y'), (x'', y'')) = d(x', x'') + d(y', y'').$$

Nous aurons pour $\varepsilon > 0$, $\limsup_{n \rightarrow +\infty} \mathbb{P}(d_m((X_n, Y_n), (X, Y)) > \varepsilon)$ est

$$\begin{aligned} &= \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, Y_n) + d(X, Y) > \varepsilon) \\ &\leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \varepsilon/2) + \mathbb{P}(d(Y_n, Y) > \varepsilon/2) \\ &\leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \varepsilon/2) + \limsup_{n \rightarrow +\infty} \mathbb{P}(d(Y_n, Y) > \varepsilon/2) \\ &= 0. \end{aligned}$$

6.3. Théorème de Skorohod-Wichura. Pour simplifier, mettons nous dans un espace métrique complet séparable. Énonçons le théorème.

THÉORÈME 5. Soit une suite de variables aléatoires $(X_n)_{n \geq 0}$ et X une autre variable aléatoire, toutes à valeurs dans (S, d) complet séparable, pas nécessairement définies sur le même espace de probabilité.

Si $X_n \rightsquigarrow X$, alors on peut construire un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ portant des variables aléatoires $(Y_n)_{n \geq 0}$ et Y , telles que

$$\mathbb{P}_X = \mathbb{P}_Y \text{ et } (\forall n \geq 0, \mathbb{P}_{X_n} = \mathbb{P}_{Y_n})$$

et

$$Y_n \rightarrow Y, \text{ p.s.}$$

Ce théorème est puissant et peut se révéler utile dans certaines situations.

7. Annexe

7.1. Intervalles F -continus où F est une fonction de distribution. Soit \mathbb{P} une probabilité \mathbb{P} sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Considérons sa fonction de répartition

$$(x_1, \dots, x_k) \mapsto F(x_1, \dots, x_k) = P\left(\prod_{i=1}^k]-\infty, x_i]\right).$$

7.1.1. *Intervalles de continuité de F .*

Soit un intervalle borné de \mathbb{R}^k de la forme

$$]a, b] = \prod_{i=1}^k]a_i, b_i].$$

Définissons

$$E(a, b) = \{c = (c_1, \dots, c_k) \in \mathbb{R}^k, \forall 1 \leq i \leq k, (c_i = a_i \text{ ou } c_i = b_i)\}.$$

Nous dirons que l'intervalle $]a, b]$ est F -continue ssi $]a, b]$ borné et nous avons

$$\forall c \in E(a, b), \mathbb{P}(\partial] - \infty, c]) = 0.$$

Soit \mathcal{U} la classe des intervalles F -continus. Par convention, nous mettons l'ensemble vide dans \mathcal{U} puisque la condition d'appartenance ne peut être vérifiée. Nous allons voir quelques propriétés de la classe \mathcal{U} .

7.1.2. \mathcal{U} est stable par intersection finies. Soit $]a, b] = \prod_{i=1}^k]a_i, b_i] \in U$ et $]c, d] = \prod_{i=1}^k]c_i, d_i]$. Nous avons

$$]a, b] \cap]c, d] = \prod_{i=1}^k]a_i \vee c_i, b_i \wedge d_i] =]\alpha, \beta]$$

où $x \vee y$ et $x \wedge y$ désignent respectivement le maximum et minimum de x et y et, $\alpha = (a_1 \vee c_1, \dots, a_k \vee c_k)$ et $\beta = (b_1 \wedge d_1, \dots, b_k \wedge d_k)$. Si $]a, b] \cap]c, d]$ est vide, il est dans \mathcal{U} . Sinon, aucun des facteurs $]a_i \vee c_i, b_i \wedge d_i]$ n'est vide. Nous allons montrer que

$$(7.1) \quad \forall e \in E(\alpha, \beta), \partial] - \infty, e] \subset \bigcup_{z \in E(a, b) \cup E(c, d)} \partial] - \infty, z].$$

En effet, Soit $e \in E(\alpha, \beta)$. On a donc

$$e_i = a_i \vee c_i \text{ ou } b_i \wedge d_i, \quad 1 \leq i \leq k.$$

Soit $t \in \partial] - \infty, e]$. Cela veut dire que

$$(t_i \leq c_i, \quad 1 \leq i \leq k) \text{ et } (\exists i_0, t_{i_0} = e_{i_0})$$

Puisque $]\alpha, \beta]$ est dans $]a, b]$ et $]c, d]$, t vérifie

$$t_i \leq b_i \text{ et } t_i \leq d_i, \quad 1 \leq i \leq k.$$

Maintenant, on considère un i_0 tel que $t_{i_0} = c_{i_0}$. Nous avons quatre cas

$$\left\{ \begin{array}{ll} t_{i_0} = e_{i_0} = a_{i_0} \vee c_{i_0} = a_{i_0} & \implies t_{i_0} = a_{i_0} \text{ et } t_i \leq b_i, 1 \leq i \leq k \\ t_{i_0} = e_{i_0} = a_{i_0} \vee c_{i_0} = c_{i_0} & \implies t_{i_0} = c_{i_0} \text{ et } t_i \leq d_i, 1 \leq i \leq k \\ t_{i_0} = e_{i_0} = b_{i_0} \wedge d_{i_0} = b_{i_0} & \implies t_{i_0} = b_{i_0} \text{ et } t_i \leq b_i, 1 \leq i \leq k \\ t_{i_0} = e_{i_0} = b_{i_0} \wedge d_{i_0} = d_{i_0} & \implies t_{i_0} = d_{i_0} \text{ et } t_i \leq d_i, 1 \leq i \leq k \end{array} \right.$$

We conclude the following conclusions for the four lines. First line : $t \in \partial] - \infty, z_1]$ where $z_1 = (b_1, \dots, b_{i_0-1}, a_{i_0}, b_{i_0+1}, b_k) \in E(a, b)$. Second line : $t \in \partial] - \infty, z_2]$ where $z_2 = (d_1, \dots, d_{i_0-1}, c_{i_0}, d_{i_0+1}, d_k) \in E(c, d)$. Third line : $t \in \partial] - \infty, b]$ and of course $b \in E(a, b)$. Fourth line : $t \in \partial] - \infty, d]$ and of course $d \in E(c, d)$. So t is one of the $\partial] - \infty, z]$ with $z \in E(a, b) \cup E(c, d)$. So 7.1 est vraie et puisque cette union est finie et, est composée de parties mesurables de probabilité nulle nous avons

$$\forall e \in E(\alpha, \beta), P(\partial] - \infty, e]) = 0.$$

Dès lors, \mathcal{U} est stable par intersection finie. Nous avons :

LEMME 2. *Tout voisinage d'un point x contient un intervalle $]a, b]$ F -continue contenant x .*

Soit V un voisinage de x . Il existe un intervalle $]a, b[$ tel que

$$x \in \prod_{i=1}^k]a_i, b_i[$$

Soit

$$\varepsilon_0 = \min(x_i - a_i, 1 \leq i \leq k) \wedge \min(b_i - x_i, 1 \leq i \leq k).$$

Notons $\delta = (1, \dots, 1)$ le vecteur de \mathbb{R}^k dont toutes les composantes sont nulles. Nous avons $0 < \varepsilon < \varepsilon_0$,

$$]a + \varepsilon\delta, x + \varepsilon\delta] \subset]a, b[.$$

Tout point e de $E(a + \varepsilon\delta, x + \varepsilon\delta)$ est de la forme

$$t(\varepsilon) = (t_1 + \varepsilon, t_2 + \varepsilon, \dots, t_k + \varepsilon)$$

avec bien sur $t_i = a_i$ ou $t_i = x_i$. Pour un choix de $t = (t_1, \dots, t_k)$, les ensembles $\partial] - \infty, t(\varepsilon)]$ sont disjoints. Donc, en dehors d'une partie dénombrable $D(t)$ de $]0, \varepsilon_0[$ on aura

$$P(\partial] - \infty, t(\varepsilon)]) = 0$$

En dehors de l'ensemble dénombrable $D = \cup_t D(t) \subset]0, \varepsilon_0[$, (puisque D est union de 2^k ensembles dénombrables), nous pouvons choisir un ε de $]0, \varepsilon_0[$ tel que pour tout e de composantes

$$e_i = a_i + \varepsilon \text{ or } x_i + \varepsilon,$$

nous avons

$$\mathbb{P}(\partial] - \infty, t(\varepsilon)]) = 0$$

et donc

$$x \in]a + \varepsilon\delta, x + \varepsilon\delta] \subset]a, b[.$$

Appelons $\varepsilon(x)$ la valeur trouvée de ε . Nous venons de montrer que pour tout voisinage V de x , il existe $]A_x, B_x[=]a + \varepsilon(x)\delta, x + \varepsilon(x)\delta/2[$ et $]a_x, a_x] =]a + \varepsilon(x)\delta, x + \varepsilon(x)\delta] \in \mathcal{U}$ tels que

$$(7.2) \quad x \in]A_x, B_x[\subset]a_x, b_x] \subset V.$$

A partir de là, montrons que *Tout ouvert G de \mathbb{R}^k est union dénombrable d'intervalles F -continue.*

En effet, d'après (7.2), tout ouvert G peut s'écrire

$$G = \bigcup_{x \in G}]A_x, B_x[.$$

Puisque \mathbb{R}^k un espace métrique séparable, ce recouvrement ouvert se réduit à un ses sous-recouvrements dénombrables, i.e., il existe une suite $(x_j)_{j \geq 0} \subset G$ telle que

$$G = \bigcup_{j \geq 0}]A_{x_j}, B_{x_j}[.$$

On en déduit que

$$G = \bigcup_{j \geq 0}]a_{x_j}, b_{x_j}].$$

où les $]a_{x_j}, b_{x_j}]$ sont des intervalles F -continus. Nous concluons par :

PROPOSITION 18. *Soit F une fonction de distribution sur \mathbb{R}^k , $k \geq 1$. Alors, tout ouvert G dans \mathbb{R}^k est une union dénombrable d'intervalles F -continus de la form $]a, b]$ or $]a, b[$, où par définition, un intervalle (a, b) est F -continu si et seulement si pour tout*

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k) \in \{0, 1\}^k,$$

le point

$$b + \varepsilon * (a - b) = (b_1 + \varepsilon_1(a_1 - b_1), b_2 + \varepsilon_2(a_2 - b_2), \dots, b_k + \varepsilon_k(a_k - b_k))$$

est un point de continuité de F .

7.2. Fonctions semi-continues. Une fonction $f : S \mapsto \overline{\mathbb{R}}$ est continue en tout x ssi

(i) pour tout $x \in \mathbb{R}$, pour tout $\varepsilon > 0$, il existe un voisinage V de x tel que

$$y \in V \Rightarrow f(y) \in]f(x) - \varepsilon, f(x) + \varepsilon[.$$

Dans cette formule, on s'intéresse à tout l'intervalle $]f(x) - \varepsilon, f(x) + \varepsilon[$. Mais on peut s'intéresser uniquement à l'une des bornes de l'intervalle. Cela nous donne les fonctions semi-continues. Précisément, f est dite semi-continue supérieurement (noté s.c.s) ssi

(ii) pour tout $x \in \mathbb{R}$, pour tout $\varepsilon > 0$, il existe un voisinage V de x tel que

$$y \in V \Rightarrow f(y) < f(x) + \varepsilon$$

Elle est dite semi-continue inférieurement (noté s.c.i) ssi

(iii) Pour tout $x \in \mathbb{R}$, pour tout $\varepsilon > 0$, il existe un voisinage V de x tel que

$$y \in V \Rightarrow f(y) > f(x) - \varepsilon$$

Nous avons ces propriétés immédiates.

(a) Une fonction numérique est continue si et seulement si elle est à la fois s.c.s and s.c.i.

(b) Une fonction numérique f est **s.c.s** si et seulement si son oppos $-f$ is s.c.i.

Voici une caractérisation des fonctions semi-continues.

PROPOSITION 19. *Nous avons les propriétés suivantes :*

(1) Une fonction $f : S \mapsto \mathbb{R}$ est s.c.s si et seulement si l'ensemble $(f \geq c)$ est fermé pour tout nombre réel c .

(2) Une fonction $f : S \mapsto \mathbb{R}$ est s.c.s si et seulement si l'ensemble $(f \geq c)$ est fermé pour tout nombre réel c .

(3) Si f est *s.c.s* or *s.c.i*, alors elle est mesurable.

Preuve. Preuve du point (1). Commençons par le sens direct. Soit f une fonction *s.c.s* from S to \mathbb{R} . Montrons que pour tout réel c fixé, l'ensemble $(f \geq c)$ est fermé en montrant que l'ensemble $(f < c)$ est ouvert. Soit $x \in G = (f < c)^c$, c'est-à-dire que, $f(x) < c$. Posons $\varepsilon = c - f(x) > 0$. Puisque f est *s.c.s*, il existe un voisinage V de x tel que

$$y \in V \Rightarrow f(y) < f(x) + \varepsilon = c,$$

ce qui peut être écrit sous la forme

$$y \in V \Rightarrow f(y) < c,$$

ce qui signifie que $V \subseteq G^c$. Nous avons prouvé que G^c contient chacun de ses points avec un de ses voisinages. Donc G^c est ouvert. Le sens direct est prouvé.

Inversement, supposons que pour tout nombre réel c , l'ensemble $(f \geq c)$ est fermé. Soit $x \in S$ quelconque. Donc pour tout $\varepsilon > 0$, $G = (f < f(x) + \varepsilon)$ est ouvert. Mais x appartient à G , donc G le contient avec un de ses voisinages $V \in V(x)$ et donc

$$y \in V \Rightarrow y \in (f < f(x) + \varepsilon) \Rightarrow f(y) \leq f(x) + \varepsilon.$$

Donc f est semi-continue supérieurement.

Le point (2) se prouve directement à partir du point 1, en utilisant la transformation $-f$.

Le point (3) découle des critères classiques de mesurabilité pour les fonctions réelles.

7.3. Propriété caractéristique d'une famille parties mesurables disjointes.

PROPOSITION 20. Soit une famille $(B_\lambda)_{\lambda \in \Gamma}$ de parties mesurables deux à deux disjointes d'un espace probabilisé (Ω, \mathcal{A}, L) . Alors un nombre au plus dénombrable d'entre elles ont une probabilité non nulle.

Preuve.

Soit l'ensemble des indices λ pour lesquels B_λ est de probabilité non nulle,

$$D = \{\lambda \in \Gamma, L(B_\lambda) > 0\}.$$

On a surement

$$D = \cup_{k \geq 1} D_k,$$

avec

$$D_k = \{\lambda \in \Gamma, L(B_\lambda) > 1/k\}.$$

Maintenant soit r éléments de D_k notés $\lambda_1, \lambda_2, \dots, \lambda_r$, on a, par le fait que les ensembles sont disjoints deux à deux,

$$1 \geq L\left(\bigcup_1^r B_{\lambda_j}\right) = \sum_1^r L(B_{\lambda_j}) \geq r/k.$$

D'où

$$r \leq k.$$

Donc D_k contient au plus k éléments. Par suite, D est une union dénombrable d'ensembles finis. D est donc au plus dénombrable.

7.4. Mesurabilité de l'ensemble des points de discontinuité.

Voilà un résultat surprenant, à savoir que l'ensemble des points de discontinuité d'une application quelconque g , noté $discont(g)$, d'un espace métrique (S, d) dans un autre (D, r) est mesurable. En effet, nous avons

LEMME 3. *Soit g une fonction de l'espace métrique (S, d) vers l'espace métrique (D, r) . Soit $discont(g)$, l'ensemble des points de continuité de g . Nous avons*

$$(7.3) \quad discont(g) = \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}$$

où pour chaque couple d'entiers (s, t) , l'ensemble

$$B_{s,t} = \{x \in S, \exists (y, z) \in S^2, d(x, y) < 1/t, d(z, x) < 1/t, r(g(y), g(z)) \geq 1/s\}.$$

est ouvert.

La conséquence de ce lemme est que $discont(g)$ est mesurable.

Preuve du lemme. Montrons d'abord que

$$\bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t} \subseteq \text{discont}(g)$$

Soit $x \in \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}$. Il existe un entier $s \geq 1 \geq \text{fixé}$ tel que pour chaque entier $t \geq 1$, il existe y_t et z_t tel tel que

$$d(x, y_t) < 1/t,$$

et

$$d(x, z_t) < 1/t,$$

et

$$(7.4) \quad \forall t \geq 1, \quad r(g(y_t), g(z_t)) \geq 1/s$$

Et si g est continue en x , alors par continuité,

$$r(g(y_t), g(z_t)) \leq r(g(y_t), g(x)) + r(g(x), g(z_t)) \rightarrow 0$$

ce qui est en contradiction avec (7.4). Donc x n'est pas un point de continuité, d'où $x \in \text{discont}(g)$.

Montrons

$$\text{discont}(g) \subseteq \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}.$$

Soit x un point de discontinuité de g . Par négation de la continuité,

$$\exists \varepsilon > 0, \forall \eta > 0, \exists y \in S, \quad d(x, y) < \eta, \quad r(g(y), g(x)) \geq \varepsilon.$$

Soit s un entier tel que $\varepsilon \geq 1/s$, donc pour tout $1/t$ où t en entier positif non nul,

$$\exists y \in S, \quad d(x, y) < 1/t, \quad r(g(y), g(x)) \geq 1/s.$$

En posant $z = x$, on a bien

$$d(x, z) < 1/t, \quad d(x, y) < 1/t, \quad r(g(y), g(x)) \geq 1/s.$$

Donc $x \in \bigcup_{s=1}^{\infty} \bigcap_{t=1}^{\infty} B_{s,t}$. D'où l'égalité.

Montrons enfin que chaque $B_{s,t}$ est ouvert. Posons $a = 1/s > 0$ et $b = 1/t > 0$. Soit $x \in B_{s,t}$, Donc

$$\exists (y, z) \in S^2, d(x, y) < b, d(z, x) < b, r(g(y), g(z)) \geq a$$

Soit $c = \min(b - d(x, y), b - d(z, x)) > 0$. Soit $x' \in B(x, c)$, donc

$$d(x', y) \leq d(x', x) + d(x, y) < c + d(x, y) \leq b$$

et

$$d(x', z) < d(x', x) + d(x, z) \leq c + d(x, z) \leq b$$

et

$$r(g(y), g(z)) \geq a$$

D'où $x' \in B_{s,t}$. D'où

$$x \in B(x, c) \subseteq B_{s,t}$$

Ainsi chaque $B_{s,t}$ contient ses points avec des boules ouvertes. Donc chaque $B_{s,t}$ est ouvert. Donc $\text{discont}(g)$ est mesurable.

7.5. Théorème de Stone-Weierstrass. cv.subsec.annexe.SW1

PROPOSITION 21. *cv.subsec.annexe.SW.prop* Soit (S, d) un espace métrique compact et H une partie non vide de $C(S, \mathbb{R})$ l'ensemble des fonctions continues de S dans \mathbb{R} vérifiant les propriétés

(i) H est réticulée, i.e., si f et g sont deux éléments de H , alors $f \wedge g$ et $f \vee g$ appartient à H

(ii) Si x et y sont éléments de S , et (a, b) un couple de réels (avec $a \neq b$ si $x \neq y$), alors il existe deux éléments h et k de H tels que

$$h(x) = a \text{ et } k(y) = b.$$

Alors H est dense de $C(S, \mathbb{R})$ munie de sa topologie uniforme, c'est-à-dire, que toute fonction continue de S dans \mathbb{R} est limite uniforme d'une suite d'éléments de H .

THÉORÈME 6. *Soit (S, d) un espace métrique compact et H une partie non vide de $C(S, \mathbb{R})$ l'ensemble des fonctions continues de S dans \mathbb{R} vérifiant les propriétés :*

(i) *H contient les fonctions constantes.*

(ii) *Si $(h, k) \in H^2$, $h + k \in H$, $h \times k \in H$, $\bar{u} \in H$.*

(iii) *H sépare les points de S , i.e., pour tous éléments x et y distinctes de S , $x \neq y$, alors il existe $h \in H$*

$$h(x) \neq h(y).$$

Alors H est dense de $C(S, \mathbb{R})$ muni de sa topologie uniforme, c'est-à-dire, que toute fonction continue de S dans \mathbb{R} est limite uniforme d'une suite d'éléments de H .

Remarque.

Si on travaille sur \mathbb{R} , le théorème est vraie et la condition $\bar{u} \in H$, n'a pas de sens.

7.6. Divers. Une relation utile. A prouver pour des réels x, y, X , et Y ,

$$(7.5) \quad |\min(x, y) - \min(X, Y)| \leq |x - X| + |y - Y|$$

En effet si $\min(x, y) = x$ et $\min(X, Y) = X$,

$$|\min(x, y) - \min(X, Y)| \leq |x - X|$$

si $\min(x, y) = y$ et $\min(X, Y) = Y$,

$$|\min(x, y) - \min(X, Y)| \leq |y - Y|$$

Maintenant soit $\min(x, y) = x$ et $\min(X, Y) = Y$. On peut supposer que $x \leq Y$. On aura

$$\min(x, y) - \min(X, Y) = Y - x \leq X - x$$

puisque $X \geq Y$. Le cas $\min(x, y) = y$ et $\min(X, Y) = X$ se traite comme le cas précédent. Donc (7.5) est vraie.

Tension uniforme et tension asymptotique

1. Introduction

Toute théorie limite possède une partie qui traite de la notion de compacité, c'est-à-dire de l'existence de sous-suites convergentes, l'équivalent du théorème de Bolzano-Weierstrass pour les suites numériques. Il s'agira pour la convergence vague de la partie sur la tension des suites de variables aléatoires. Au plan général, le théorème de Prohorov prévaut et affirme que toute suite asymptotiquement tendue de variables aléatoires possède une sous-suite vaguement convergente.

Dans ce chapitre, nous allons nous restreindre au cas où S est \mathbb{R}^k , puisque le traitement entre le cas général et le cas de \mathbb{R}^k sont un peu éloignés l'un de l'autre. Dans ce cas particulier, le théorème de Helly-Bray joue le grand rôle.

Ce chapitre est donc centré sur \mathbb{R}^k , où $k \geq 1$. Deux points sont derrière toutes les propriétés exposées ici. Le premier est que les parties compactes de \mathbb{R}^k , sont les ensembles de fermés et bornés. Le second est que \mathbb{R}^k est un espace métrique complet séparable.

Dans ce chapitre, sauf indication contraire, la norme *max* ainsi définie

$$\|x\| = \max_{1 \leq i \leq k} |x_i|.$$

est utilisée. Ainsi les boules ouvertes $B(x, r)$ et fermés $B^f(x, r)$ sont définies par

$$B(x, r) = \{x \in \mathbb{R}^k, \|x\| < r\} = \prod_{i=1}^k]x_i - r, x_i + r[$$

pour $x = (x_1, \dots, x_k)$ et $r > 0$ et

$$B^f(x, r) = \{x \in \mathbb{R}^k, \|x\| \leq r\} = \prod_{i=1}^k [x_i - r, x_i + r]$$

pour $x = (x_1, \dots, x_k)$ et $r \geq 0$.

Avant de commencer, faisons quelques notations.

Soit $a = (a_1, \dots, a_k)$ et $b = (b_1, \dots, b_k)$. Définissons ces relations d'ordre

$$(a \leq b) \iff (\forall (1 \leq i \leq k), a_i \leq b_i),$$

ensuite,

$$(a < b) \iff (\forall (1 \leq i \leq k), a_i \leq b_i, \exists (1 \leq i_0 \leq k), a_{i_0} < b_{i_0})$$

et enfin,

$$(a \prec b) \iff (\forall (1 \leq i \leq k), a_i < b_i,)$$

et son symétrique

$$(a \succ b) \iff (\forall (1 \leq i \leq k), a_i > b_i,)$$

Définissons ces classes d'ensembles compacts.

Pour $A = (A_1, \dots, A_k) \prec V = (B_1, \dots, B_k)$, notons

$$K_{A,B} = \prod_{i=1}^k [A_i, B_i].$$

Pour $A = (A_1, \dots, A_k) \succ 0$, notons

$$K_A = \prod_{i=1}^k [-A_i, A_i].$$

Pour $M \in \mathbb{R}$, $M > 0$, notons

$$K_{c,M} = [-M, M]^k.$$

Ces ensembles $K_{A,B}$, K_A et $K_{c,M}$, tous compacts, servent à définir la tension de suites de vecteurs aléatoires. Ils jouent des rôles équivalents dans la définition de la tension. Pour cela énonçons ici ce résultat d'équivalence.

PROPOSITION 22. *Soit $\{\mathbb{P}_n, n \geq 1\}$ une suite de probabilités sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Les propositions suivantes sont équivalentes.*

(1a) *Pour tout $\varepsilon > 0$, il exist un compact K de \mathbb{R}^k tel que*

$$\inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(2a) *Pour tout $\varepsilon > 0$, il existe un réel $M > 0$ tel que*

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

(3a) Pour tout $\varepsilon > 0$, il existe un vecteur $A = (A_1, \dots, A_k) \succ 0$, de \mathbb{R}^k tels que

$$\inf_{n \geq 1} \mathbb{P}_n(K_A) \geq 1 - \varepsilon.$$

(4a) Pour tout $\varepsilon > 0$, il existe deux vecteurs $A = (A_1, \dots, A_k) \prec V = (B_1, \dots, B_k)$ de \mathbb{R}^k tels que

$$\inf_{n \geq 1} \mathbb{P}_n(K_{A,B}) \geq 1 - \varepsilon.$$

(1b) Pour tout $\varepsilon > 0$, il exist un compact K de \mathbb{R}^k tel que

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(2b) Pour tout $\varepsilon > 0$, il existe un réel $M > 0$ tel que

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

(3b) Pour tout $\varepsilon > 0$, il existe un vecteur $A = (A_1, \dots, A_k) \succ 0$, de \mathbb{R}^k tels que

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K_A) \geq 1 - \varepsilon.$$

(4b) Pour tout $\varepsilon > 0$, il existe deux vecteurs $A = (A_1, \dots, A_k) \prec V = (B_1, \dots, B_k)$ de \mathbb{R}^k tels que

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K_{A,B}) \geq 1 - \varepsilon.$$

PREUVE. Il faut noter que nous avons deux groupes de formules : (1a)–(4a) et (1b)–(4b). En fait, nous allons démontrer les équivalences des différents points de chaque groupe, puis entre les deux premiers points des deux groupes.

Equivalence entre les points du groupe (1a)-(4b): Soit $\varepsilon > 0$ fixé. Montrons :

(1a) \implies (2a). Soit K tel $\sup_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon$. Puisque K est compact, il est borné. Il est dans un ensemble $\{x, \|x\| \leq M\} = K_{c,M}$ et

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq \inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(2a) \implies (3a). Ceci est évident, puisque un $K_{c,M}$ est un K_A avec $A = (M, M, \dots, M)$.

(3a) \implies (4a). Ceci est aussi évident, puisque un K_A , pour $A \succ 0$, est exactement $K_{-A,A}$.

(4a) \implies (1a). Ceci est encore évident puisque $K_{A,B}$ est un compact de \mathbb{R}^k .

Equivalence entre les points du groupe (1b)-(4b). La preuve est exactement la même que pour le premier groupe.

Equivalence entre les deux groupes. Il suffit de prouver ceci :
(1a) \iff (1b)

Si (1a) est vraie, alors pour tout $\varepsilon \geq 1$, il existe un compact K de \mathbb{R}^k tel que, pour tout $n \geq 1$,

$$\mathbb{P}_n(K) \geq 1 - \varepsilon.$$

Dès lors

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n(K) \geq 1 - \varepsilon,$$

ce qui donne (1b).

Si (1b) est vraie, alors, pour tout $\varepsilon \geq 1$, il existe un compact K tel que

$$\left\{ \sup_{n \geq 1} \inf_{p \geq n} \mathbb{P}_p(K) \right\} \geq 1 - \varepsilon/2.$$

Alors, il existe $N \geq 1$, tel que

$$\inf_{p \geq N+1} \mathbb{P}_p(K) \geq 1 - \varepsilon,$$

c'est-à-dire que pour tout $n > N$,

$$\mathbb{P}_n(K) \geq 1 - \varepsilon.$$

Puisque K est un ensemble compact, il est dans un K_{c,M_∞} et alors, pour tout $n > N$,

$$\mathbb{P}_n(K_{c,M_\infty}) \geq 1 - \varepsilon.$$

Maintenant, pour chaque $1 \leq j \leq N$ fixé, l'ensemble $(\|x\| \leq M) = K_{c,M}$ croît, avec M , vers \mathbb{R}^k et donc, $\mathbb{P}_j(\|X_j\| \leq M) \uparrow 1$. Ainsi, il existe $M_j > 0$ réel tel que pour tout $1 \leq j \leq N$,

$$\mathbb{P}_j(K_{c,M_j}) \geq 1 - \varepsilon.$$

Nous venons de montrer que toute probabilité $\mathbb{P}^{(0)}$ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est **tendue**, c'est-à-dire que pour tout $\varepsilon > 0$, il existe un compact $K^{(0)} = K_{c,M^{(0)}}$ de \mathbb{R}^k tel que

$$(1.1) \quad \mathbb{P}^{(0)}(K^{(0)}) \geq 1 - \varepsilon.$$

Maintenant, si on prend

$$M = \max(M_1, \dots, M_N, M_\infty),$$

nous avons que les K_{c,M_j} , $1 \leq j \leq M$ et K_{c,M_∞} sont tous dans $K_{c,M}$, un compact de \mathbb{R}^k et enfin pour tout $n \geq 1$,

$$\mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon$$

et enfin

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon,$$

ce qui est (1a).

Voici un autre lien entre de telles formules et les fonctions de répartition associés. A toute probabilité, nous associons sa fonction de répartition

$$F_{\mathbb{P}}(x) = \mathbb{P}([-\infty, x]) = P\left(\prod_{i=1}^k [-\infty, x_i]\right), x = (x_1, \dots, x_k) \in \mathbb{R}^k.$$

Cette fonction de répartition, à son tour, détermine la probabilité de Lebesgues-Stieljes ainsi : pour tout $(a, b) \in \mathbb{R}^k \times \mathbb{R}^k$, $a \leq b$,

$$\mathbb{P}([a, b]) = \Delta_{a,b}F = \sum_{\epsilon \in \{0,1\}^k} (-1)^{s(\epsilon)} F(b + \epsilon * (a - b)) \geq 0,$$

où pour $\epsilon = (\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k$, $s(\epsilon) = \epsilon_1 + \dots + \epsilon_k$, pour $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, $y = (y_1, \dots, y_k)$, $x * y = (x_1 y_1, \dots, x_k y_k)$.

Nous allons utiliser les mesures de Lebesgues-Stieljes que le lecteur peut réviser dans les ouvrages [8] et surtout dans le chapitre 1 de [7]. Nous avons besoin de cette notation. Soit $M > 0$. Notons

$$L_M = \{x, \exists(1 \leq i \leq k), x_i \leq -c\}$$

Nous avons :

PROPOSITION 23. *Soit $\{\mathbb{P}_n, n \geq 1\}$ une suite de probabilités sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ et soit la suite de leurs fonctions de répartition $\{F_n \geq 1\}$ avec $F_{\mathbb{P}_n} = F_n$ for $n \geq 1$. Alors les trois points suivants sont équivalentes entre eux.*

(1c) *Pour tout $\varepsilon > 0$, il existe $0 < C \in \mathbb{R}^k$ et il existe $c > 0$ such that*

$$\inf_{n \geq 1} F_n(C) \geq 1 - \varepsilon$$

and

$$\inf_{n \geq 0} \mathbb{P}_n(L_c) \leq \varepsilon.$$

(2c) Pour tout $\varepsilon > 0$, il existe $0 < c$ tel que pour $c^{(k)} = (c, \dots, c)$ et il existe $M > 0$ such that

$$\inf_{n \geq 1} F_n(c^{(k)}) \geq 1 - \varepsilon$$

and

$$\inf_{n \geq 0} \mathbb{P}_n(L_M) \leq \varepsilon.$$

(3c) Pour tout $\varepsilon > 0$, il existe $M > 0$ tel que

$$\inf_{n \geq 1} P_n(K_{c,M}) \geq 1 - \varepsilon.$$

Puisque le point (3c) est le point (2c) de la proposition 22, alors les points (3a) et (3b) sont équivalents à tous les points de cette proposition.

Preuve. Procédons aux preuves des différentes équivalences.

Preuve de (1c) \implies (2c). Pour tout $\varepsilon > 0$, il existe $0 < C \in \mathbb{R}^k$ such that

$$\inf_{n \geq 1} F_n(C) \geq 1 - \varepsilon.$$

Soit $c = \max\{C_i, 1 \leq i \leq k\}$. we have $]-\infty, C] \subset]-\infty, c^{(k)}]$ and $F_n(c^{(k)}) \geq F_n(C)$,

$$\inf_{n \geq 1} F_n(c^{(k)}) \geq 1 - \varepsilon.$$

La preuve est terminée puisque la deuxième formule est la même pour les deux points.

Preuve de (2c) \implies (3c). De (2c), nous tirons un $d^{(k)} = (d, \dots, d)$, avec $d > 0$, tel que

$$\inf_{n \geq 1} F_n(d^{(k)}) \geq 1 - \varepsilon/2,$$

et un nombre $e > 0$ tel que

$$\inf_{n \geq 1} \mathbb{P}_n(L_e) \leq \varepsilon/2.$$

En faisant $M = \max(d, e)$, nous avons

$$\inf_{n \geq 1} F_n(M^{(k)}) \geq 1 - \varepsilon/2.$$

et un nombre $e > 0$ tel que

$$\inf_{n \geq 1} \mathbb{P}_n(L_M) \leq \varepsilon/2.$$

Maintenant, décomposons \mathbb{R}^k en $\mathbb{R}^k = L_M + L_M^c$, avec

$$L_M^c = \{x, \forall (1 \leq i \leq k), x_i \geq -M\}$$

qui se décompose lui-même en

$$\begin{aligned} L_M^c &= \{x, \forall (1 \leq i \leq k), -M \leq x_i \leq M\} \\ &+ \{x, \forall (1 \leq i \leq k), x_i \geq -M \text{ et } \exists (1 \leq i \leq k), x_i > M\} \\ &= K_{c,M} + B, \end{aligned}$$

où, de manière évidente,

$$B \subset]-\infty, M^{(k)}]^c.$$

Dès lors, on peut tirer du fait que $\mathbb{R}^k = L_M + K_{c,M} + B$, que

$$(1.2) \quad K_{c,M}^c = L_M + B.$$

Donc, pour tout $n \geq 1$,

$$\mathbb{P}_n(K_{c,M}) = \mathbb{P}_n(L_M) + \mathbb{P}_n(B) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

puisque $B \subset]-\infty, M^{(k)}]^c$ et donc pour tout $n \geq 1$,

$$\begin{aligned} \mathbb{P}_n(B) &\leq \mathbb{P}_n(]-\infty, M^{(k)}]^c) \\ &\leq 1 - \mathbb{P}_n(]-\infty, M^{(k)}]) \\ &\leq 1 - F_n(M^{(k)}) \leq \varepsilon/2. \end{aligned}$$

Ce qui termine la preuve de ce point.

Preuve de (3c) \implies (1c). Supposons (3c) : pour tout $\varepsilon > 0$, il existe $M > 0$ tel que

$$\inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

Nous avons

$$\inf_{n \geq 1} F_n(M^{(k)}) = \inf_{n \geq 1} \mathbb{P}_n(]-\infty, M^{(k)}]) \geq \inf_{n \geq 1} \mathbb{P}_n(K_{c,M}) \geq 1 - \varepsilon.$$

Ensuite, à cause de (1.2),

$$\mathbb{P}_n(L_M) \leq \mathbb{P}_n(K_{c,M}^c) \leq \varepsilon.$$

Donc (1c) a lieu. Et la boucle est bouclée.

Maintenant, exposons la notion de tension.

2. Tension

2.1. Tension individuelle. Dans ce cas particulier, toute probabilité sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est tendue dans le sens suivant

PROPOSITION 24. *Pour toute probabilité \mathbb{P} sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est tendue, c'est-à-dire que pour tout $\varepsilon > 0$, il existe un compact de \mathbb{R}^k tel que*

$$\mathbb{P}(K) \geq 1 - \varepsilon.$$

Ce résultat est déjà montré dans (1.1). L'enjeu se trouve dans la tension uniforme ou tension asymptotique.

2.2. Tension asymptotique. Tension uniforme.

DEFINITION 2. (a) *Une suite de probabilités $\{\mathbb{P}_n, n \geq 1\}$ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est dite asymptotiquement tendue ou uniformément tendue ssi pour tout $\varepsilon > 0$, il existe un compact K de \mathbb{R}^k tel que*

$$(2.1) \quad \inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \varepsilon$$

ou de manière équivalente

$$(2.2) \quad \liminf_{n \rightarrow \infty} \mathbb{P}_n(K) \geq 1 - \varepsilon.$$

(b) *Une suite de variables aléatoires $\{X_n, n \geq 1\}$ à valeurs dans $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est asymptotiquement tendue ou uniformément tendue ssi la famille de probabilités $\{\mathbb{P}_{X_n}, n \geq 1\}$ est asymptotiquement tendue ou uniformément tendue, c'est-à-dire, pour tout $\varepsilon > 0$, il existe un compact K de \mathbb{R}^k tel que*

$$\inf_{n \geq 1} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon$$

ou de manière équivalente

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon.$$

(c) *Une suite $\{F_n, n \geq 1\}$ de fonctions de répartition sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est asymptotiquement tendue ou uniformément tendue ssi la suite des mesures de Lebesgues-Stieljes $\{\mathbb{P}_n, n \geq 1\}$ définie par ses éléments est*

asymptotiquement tendue ou uniformément tendue, c'est-à-dire, pour tout $\varepsilon > 0$, il existe un compact K de \mathbb{R}^k tel que

$$\inf_{n \geq 1} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon$$

ou de manière équivalente, si et seulement si, pour tout $\varepsilon > 0$, il existe $0 < C \in \mathbb{R}^k$ et il existe $c > 0$ tels que

$$\inf_{n \geq 1} F_n(C) \geq 1 - \varepsilon$$

et

$$\inf_{n \geq 0} \mathbb{P}_n(L_c) \leq \varepsilon,$$

si et seulement si, pour tout $\varepsilon > 0$, il existe $0 < c$ tel que pour $c^{(k)} = (c, \dots, c)$,

$$\inf_{n \geq 1} F_n(c^{(k)}) \geq 1 - \varepsilon$$

et

$$\inf_{n \geq 0} \mathbb{P}_n(L_c) \leq \varepsilon,$$

Il vient de la proposition 22 que l'équivalence entre la tension uniforme (2.1) et la tension (2.2) sont identiques dans le cas spécifique de \mathbb{R}^k . Pour cette raison, nous parlerons uniquement de famille tendue de mesure de probabilités ou de variables aléatoires.

Avant d'en venir au théorème de Helly-Bray, donnons trois importantes propriétés de la tension.

2.3. Tension et transformation continue. La tension d'une famille de variables aléatoires est conservée par transformation continue. Nous avons :

PROPOSITION 25. *Soit une suite de variables aléatoires $\{X_n, n \geq 1\}$ à valeurs dans $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ et $g : \mathbb{R}^k \mapsto \mathbb{R}^m$, $m \geq 1$, une fonction continue. Alors, si $\{X_n, n \geq 1\}$ est tendue, alors $\{g(X_n), n \geq 1\}$ est tendue.*

Preuve. Soit $\{X_n, n \geq 1\}$ tendue et $g : \mathbb{R}^k \mapsto \mathbb{R}^m$. Pour tout $\varepsilon > 0$,

il existe un compact K de \mathbb{R}^k tel que

$$(2.3) \quad \inf_{n \geq 1} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon.$$

Mais $(X_n \in K) \subset (g(X_n) \in g(K))$, où

$$K_0 = g(K) = \{g(x), x \in K\}$$

est l'image directe de K par g , qui est un ensemble compact. En effet, toute suite de K_0 possède une sous-suite convergente dans K_0 . Pour le voir soit $\{g(x_n), x_n \in K, n \geq 1\}$ une suite dans K_0 . Puisque K est compact, la suite $(x_n)_{n \geq 1}$, qui est dans K , possède une sous-suite convergente dans K , soit $x_{n(k)} \rightarrow x \in K$. Puisque g est continue, $g(x_{n(k)}) \rightarrow g(x) \in K_0$. Il s'en suit que K_0 est un compact de \mathbb{R}^m et

$$\inf_{n \geq 1} \mathbb{P}(g(X_n) \in K_0) \geq \inf_{n \geq 1} \mathbb{P}(X_n \in K) \geq 1 - \varepsilon.$$

Ce qui finit la preuve.

2.4. Caractérisation de la tension de vecteurs par celles des composantes. Dans le cas particulier de \mathbb{R}^k , nous avons cette caractérisation :

PROPOSITION 26. *Soit une suite de variables aléatoires $\{X_n, n \geq 1\}$ à valeurs dans $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$. Alors $\{X_n, n \geq 1\}$ est tendue si et seulement les suites des composantes $\{X_n^{(i)}, n \geq 1\}$, $1 \leq i \leq k$, sont tendues.*

Preuve. Soit une suite de variables aléatoires $\{X_n, n \geq 1\}$ à valeurs dans $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$.

Supposons qu'elle est tendue. Par la proposition 25, chaque suite de composantes $\{X_n^{(i)}, n \geq 1\} = \{\pi_i(X_n), n \geq 1\}$, $1 \leq i \leq k$, est une transformation continue par la i -ième projection π_i . Elle est donc tendue.

Supposons que chaque de composante $\{X_n^{(i)}, n \geq 1\}$, $1 \leq i \leq k$, est tendue. Donc, pour tout $\varepsilon > 0$, pour tout $1 \leq i \leq k$, il existe $A_i > 0$ such

$$\inf_{n \geq 1} \mathbb{P}(X_n^{(i)} \in [-A_i, A_i]) \geq 1 - \varepsilon/k.$$

En posant $A = (A_1, \dots, A_k)$, nous avons $A > 0$ et puisque

$$\bigcap_{i=1}^k (X_n^{(i)} \in [-A_i, A_i]) = \left(X_n \in \prod_{i=1}^k [-A_i, A_i] \right),$$

il en résulte que, pour tout $n \geq 1$,

$$\begin{aligned} \mathbb{P} \left(X_n \notin \prod_{i=1}^k [-A_i, A_i] \right) &= \mathbb{P} \left(\bigcup_{i=1}^k (X_n^{(i)} \notin [-A_i, A_i]) \right) \\ &\leq \sum_{i=1}^k P(X_n^{(i)} \notin [-A_i, A_i]) \leq \varepsilon, \end{aligned}$$

si bien que pour tout $n \geq 1$,

$$\mathbb{P}(X_n \in K_A) \geq 1 - \varepsilon.$$

La suite $\{X_n, n \geq 1\}$ est tendue. La preuve est achevée.

2.5. Une suite convergente vaguement est tendue. Nous avons ce résultat intéressant qui fonde la théorie de la tension. Il repose sur la tension de la limite, qui est une probabilité sur \mathbb{R}^k , donc tendue par la proposition 24.

PROPOSITION 27. *Soit une suite de variables aléatoires $\{X_n, n \geq 1\}$ à valeurs dans $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$ convergente vaguement vers X . Alors elle est tendue.*

Preuve. Utilisons la tension de \mathbb{P}_X . Pour tout $\varepsilon > 0$, il existe un compact $K_A = [-A, A]$ de \mathbb{R}^k tel que

$$\mathbb{P}(X \in K) \geq 1 - \varepsilon.$$

Soit $0 < \delta < 1$ et pose $A + \delta = (A_1 + \delta, \dots, A_k + \delta)$ et

$$\overset{\circ}{K}_{A+\delta} = \prod_{i=1}^k]-A_i - \delta, A_i + \delta[.$$

Puisque $X_n \rightsquigarrow X$ et que $\overset{\circ}{K}_{A+\delta}$ est ouvert, nous pouvons utiliser le point (ii) du théorème Portmanteau pour tout $0 < \delta < 1$

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K_{A+1}) \geq \liminf_{n \rightarrow \infty} P(X_n \in \overset{\circ}{K}_{A+\delta}) \geq \mathbb{P}(X \in \overset{\circ}{K}_{A+\delta}),$$

donc pour tout $0 < \delta < 1$,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K_{A+1}) \geq \mathbb{P}(X \in \overset{\circ}{K}_{A+\delta}),$$

Puisque K est fermé, $\overset{\circ}{K}_{A+\delta} \downarrow \overline{K} = K$ quand $\delta \downarrow 0$ et alors

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in K_{A+1}) \geq \mathbb{P}(X \in K) \geq 1 - \varepsilon.$$

Il s'en suit que la suite $\{X_n, n \geq 1\}$ est asymptotiquement tendue. Elle a en fait hérité la tension de X .

Enfin, nous pouvons maintenant aborder le théorème fondamental de la tension

3. Théorème de compacité de Prohorov dans \mathbb{R}^k .

Le théorème suivant est l'inverse de la proposition 27, du point de vue de la convergence de sous-suites.

THÉORÈME 7. (*Prohorov - Helly-Bray*) *Soit une suite tendue de variables aléatoires $\{X_n, n \geq 1\}$ à valeurs dans $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$. Alors la suite $\{X_n, n \geq 1\}$ contient une sous-suite $\{X_{n(k)}, k \geq 1\}$ convergente vers une variable aléatoire de probabilité $L = \mathbb{P}_X$.*

Ce théorème peut être directement démontré comme fait dans les ouvrages de Billingsley [2] et van der Vaart et Wellner [10]. Dans ce texte, nous allons passer par le théorème de Helly-Bray comme dans les ouvrages de van der Vaart [11] et Loève [9].

Mais, nous passons par le théorème de Helly-Bray ci-dessous.

THÉORÈME 8. (*Helly-Bray*) *Toute suite tendue $\{F_n, n \geq 1\}$ de fonctions de répartition définies sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ contient une sous-suite $\{F_{n(k)}, k \geq 1\}$ convergente vaguement vers une fonction de distribution F qui n'est pas nécessairement une fonction de répartition.*

Preuve. Soit une suite $\{F_n, n \geq 1\}$ de fonctions de répartition définies sur $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$. Notons \mathbb{Q}^k l'ensemble des éléments de \mathbb{R}^k à composantes rationnelles. \mathbb{Q}^k est dénombrable et dense dans \mathbb{R}^k . Enumérons \mathbb{Q}^k sous la forme $\mathbb{Q}^k = \{q_1, q_2, \dots\}$. Nous allons maintenant procéder à plusieurs étapes.

Etape 1. Trouvons F sur \mathbb{Q}^k . Nous allons mettre en place la construction classique de la suite diagonale. Nous avons que $(F_n(q_1))_{n \geq 1} \subset [0, 1]$. D'après la propriété de Bolzano-Weierstrass sur \mathbb{R} , cette suite admet une sous-suite $(F_{1,n}(q_1))_{n \geq 1}$ convergente vers $F(q_1)$.

Ensuite, nous appliquons la sous-suite de fonctions de répartition $(F_{1,n})_{n \geq 1}$ à q_2 ainsi : $(F_{1,n}(q_2))_{n \geq 1} \subset [0, 1]$. Nous en déduisons une sous-suite

$(F_{2,n}(q_2))_{n \geq 1}$ convergente vers $F(q_2)$. Nous précédon ainsi de proche en proche. Nous obtenons des sous-suites $(F_{j,n})_{n \geq 1}$, $j = 1, 2, \dots$ vérifiant

(a) Pour tout $j \geq 1$, $(F_{j+1,n})_{n \geq 1}$ est une sous-suite de chaque $(F_{i,n})_{n \geq 1}$ $1 \leq i \leq j$.

(b) Pour tout $j \geq 1$, pour tout $1 \leq j \leq i$, $F_{j,n}(q_i) \rightarrow F(q_i)$.

Maintenant, considérons la suite diagonale $(F_{j,j})_{j \geq 1}$. Vous pouvez voir avec l'aide d'un tableau simple, comme celui ci-dessous que : pour tout $i \geq 1$ fixé, la suite $\{F_{j,j}, j \geq i\}$ est une sous-suite de $(F_{i,n})_{n \geq i}$ et donc

$$F_{j,j}(q_i) \rightarrow F(q_i).$$

Pour lire ce tableau, il faut retenir que chaque ligne est une sous-suite des celles qui la devancent. On voit ainsi que chaque terme diagonal $F_{j,j}$ est membre de toutes les lignes de 1 à j .

$\mathbf{F}_{1,1}$	$F_{1,2}$	$F_{1,3}$	$F_{1,4}$	$F_{1,5}$	$F_{1,6}$	$F_{1,7}$	$F_{1,8}$	$F_{1,9}$	$F_{1,10}$	\dots
	$\mathbf{F}_{2,2}$	$F_{2,3}$	$F_{2,4}$	$F_{2,5}$	$F_{2,6}$	$F_{2,7}$	$F_{2,8}$	$F_{2,9}$	$F_{2,10}$	\dots
		$\mathbf{F}_{3,3}$	$F_{3,4}$	$F_{3,5}$	$F_{3,6}$	$F_{3,7}$	$F_{3,8}$	$F_{3,9}$	$F_{3,10}$	\dots
			$\mathbf{F}_{4,4}$	$F_{4,5}$	$F_{4,6}$	$F_{4,7}$	$F_{4,8}$	$F_{4,9}$	$F_{4,10}$	\dots
				\dots	\dots	\dots	\dots	\dots	\dots	\dots
					$\mathbf{F}_{j,j}$	$F_{j,j+1}$	$F_{j,j+2}$	$F_{j,j+3}$	$F_{j,j+5}$	\dots

Nous concluons que la sous-suite diagonale $(F_{j,j})_{j \geq 1}$ renommée $(F_{n(j)})_{j \geq 1}$ vérifie

$$\forall q \in \mathbb{Q}^k, F_{n(j)}(q) \rightarrow F(q) \text{ quand } j \rightarrow +\infty.$$

Etape 2. Propriétés de F sur \mathbb{Q}^k . Pour tout $(a, b) \in \mathbb{Q}^k \times \mathbb{Q}^k$, quand $j \rightarrow +\infty$,

$$\begin{aligned} 0 \leq \Delta_{a,b} F_{n(j)} &= \sum_{\epsilon \in \{0,1\}^k} (-1)^{s(\epsilon)} F_{n(j)}(b + \epsilon * (a - b)) \\ &\rightarrow \Delta_{a,b} F = \sum_{\epsilon \in \{0,1\}^k} (-1)^{s(\epsilon)} F(b + \epsilon * (a - b)) \geq 0, \end{aligned}$$

puisque tous les points $b + \epsilon * (a - b)$ sont dans \mathbb{Q}^k . Il s'en suit que F est à volume positif sur \mathbb{Q}^k .

F est aussi croissant sur \mathbb{Q}^k en héritant la croissance des $F_{n(j)}$, $j \geq 1$, on \mathbb{Q}^k .

Etape 3. Définissons G à sur $\mathbb{J}^k = \mathbb{R}^k \setminus \mathbb{Q}^k$ par la formule

$$G(x) = \inf\{F(q), q \in \mathbb{Q}^k, x < q\} \in [0, 1].$$

pour $x \in \mathbb{J}^k$. Il est évident que G est bien définie maintenant sur \mathbb{J}^k . Elle est aussi évidemment croissante au sens large.

(a) Montrons que G est continue à droite. Soit $x \in \mathbb{J}^k$ et soit $\varepsilon > 0$. Par définition de l'infimum, il existe $q \in \mathbb{Q}^k$ tel que $F(q) < G(x) + \varepsilon$. Pour tout $y \in \mathbb{J}^k$, $x < y < q$, $F(y) \leq G(q)$ and $\varepsilon > F(q) - G(x) \geq G(y) - G(x)$. Donc

$$(3.1) \quad \forall \varepsilon > 0, \exists q, x < y < q \implies 0 \leq G(y) - G(x) < \varepsilon.$$

Alors G est continue à droite.

(c) Montrons que $F_{n(j)}(x) \rightarrow G(x)$ aux points de continuité x de G .

Soit x un point de continuité de G . Pour tout $\varepsilon > 0$, nous pouvons trouver $(y', y'') \in \mathbb{Q}^k$ tel que $y' < x < y''$ et $G(y'') - G(y') < \varepsilon/2$. Soit $(y', y'') \in \mathbb{Q}^k$ tel que $y' < q' < x < q'' < y''$. Then $F(q'') - F(q') \leq G(y'') - G(y') \leq \varepsilon$. Next

$$\begin{aligned} G(y') \leq F(q') &= F_{n(j)}(q') \leq \liminf_{j \rightarrow +\infty} F_{n(j)}(x) \\ &\leq \limsup_{j \rightarrow +\infty} F_{n(j)}(x) \\ &\leq F_{n(j)}(q'') = F(q'') \leq G(y''). \end{aligned}$$

Ainsi les nombres

$$\liminf_{j \rightarrow +\infty} F_{n(j)}(x)$$

et

$$\limsup_{j \rightarrow +\infty} F_{n(j)}(x)$$

et $F(x)$ sont tous dans l'intervalle $[G(y'), G(y'')]$ de longueur égale au plus à ε . Cela qui implique

$$(3.2) \quad \max\left(\left|G(x) - \liminf_{j \rightarrow +\infty} F_{n(j)}(x)\right|, \left|G(x) - \limsup_{j \rightarrow +\infty} F_{n(j)}(x)\right|\right) \leq \varepsilon,$$

pour tout $\varepsilon > 0$. Dès lors,

$$F_{n(j)}(x) \rightarrow G(x) \text{ as } j \rightarrow +\infty.$$

(c) Montrons que G est à volume positif.

Pour tout $(a, b) \in \mathbb{J}^k \times \mathbb{J}^k$, soit $q' \downarrow a$ et $q'' \downarrow b$ avec $q' > a$ and $q'' > b$ and $(q', q'') \in \mathbb{Q}^k \times \mathbb{Q}^k$. Par limite décroissante et par définition de G ,

$$(3.3) \quad 0 \leq \Delta_{q', q''} F \rightarrow \Delta_{a, b} G \geq 0.$$

Résumé. G est une fonction de distribution sur \mathbb{J}^k et $F_{n(j)}(x) \rightarrow G(x)$

Etape 4 (finale). Maintenant, on peut prolonger G sur \mathbb{Q}^k par

$$G(q) = \inf\{F(x), x \in \mathbb{J}^k, q < x\} \in [0, 1], \quad q \in \mathbb{Q}^k.$$

En retour montrons que G est continue à droite. Que G soit continue à droite en $x \in \mathbb{J}^k$ vient de (3.1)

$$\forall \varepsilon > 0, \exists q, x < y < q \implies 0 \leq G(y) - G(x) < \varepsilon.$$

Cette formule est vraie, à l'origine, pour $y \in \mathbb{J}^k$. Pour l'étendre, choisissons un $y_0 \in \mathbb{J}^k$ et soit $q_0 \in \mathbb{Q}^k$ tel que

$$x < y_0 < q_0 < q.$$

Alors pour tout $z \in \mathbb{R}^k$

$$x < z < y_0 \implies G(z) - G(x) < \varepsilon.$$

Puisque si $z \in \mathbb{Q}^k$, nous avons

$$x < z < y_0 \implies G(z) - G(x) \leq G(y) - G(x) \leq \varepsilon.$$

Montrons que G soit continue à droite en $x \in \mathbb{Q}^k$. Re-utilisons la technique qui a donné (3.1). Par définition, pour tout $\varepsilon > 0$, il existe $y \in \mathbb{J}^k$ tel que $G(x) \leq G(y) < G(x) + \varepsilon$ avec $x < y$. Nous pouvons trouver $q_0 \in \mathbb{J}^k$ tel que $x < q_0 < y$. A partir de là, nous reconduisons es mêmes méthodes pour avoir

$$x < z < q_0 \implies G(z) - G(x) \leq \varepsilon.$$

Maintenant, en utilisant la continuité à droite et le fait que G est à volume positif sur \mathbb{Q}^k pour l'étendre à \mathbb{R}^k , en procédant de la manière qui a conduit à (3.3).

Enfin, pour démontrer que $F_{n(j)}(x) \rightarrow G(x)$ aux points de continuité, il suffit de le faire pour $x \in \mathbb{Q}^k$. Il suffit de reconsidérer la technique qui a conduit à (3.2) et de remplacer $(y', y'') \in \mathbb{Q}^k$ par $(z', z'') \in \mathbb{J}^k$ points de continuité de G avec $z' \uparrow x$ and $z'' \downarrow x$.

Pouvoir choisir de tels points de continuité vient du fait que les frontières des ensembles $] - \infty, z']$ et $] - \infty, z'']$ sont distincts lorsque z' croît strictement et z'' décroît strictement. On peut donc les choisir tels que leur valeur par la mesure $m_G([a, b]) = \Delta_{a,b}G$ soient nulles, ce qui fait des z' and z'' des points de continuité. Nous pensons que le lecteur est déjà familier avec ce type de raisonnement depuis l'exposé sur la mesure de l'exposition.

Remarque sur la preuve. Nous avons voulu avoir une preuve aussi complète que possible. L'étape 4 est superflue si on peut montrer dès l'étape 2 que la fonction F est continue à droite sur \mathbb{Q}^k . En effet la preuve dans van der Vaart [11] ne permet pas à notre avis, de dire que F est continue en $x \in \mathbb{Q}^k$.

Maintenant passons à la preuve du théorème de Prohorov.

Preuve du théorème 7 de Prohorov. Supposons que la suite de fonctions de répartition $\{F_n, n \geq 1\}$ soit tendue, c'est-à-dire que la suite des mesures de Lebesgues-Stieljes $\{\mathbb{P}_n([a, b]) = \Delta_{a,b}F_n, n \geq 1\}$ est tendue. D'après la proposition 23, pour tout $\varepsilon > 0$, nous avons un $C > 0$, $C \in \mathbb{R}^k$ tel que pour tout $n \geq 1$,

$$F_n(C) \geq 1 - \varepsilon.$$

D'après le théorème 8, il existe une sous-suite $(F_{n(j)})_{j \geq 1}$ de $(F_n)_{n \geq 1}$ convergente vaguement vers une fonction de distribution F associée à une mesure L définie par $L([a, b]) = \Delta_{a,b}F$ et bornée par l'unité.

Considérons la famille $\{C_h = C + h^{(k)}, \text{ avec } h > 0\}$. Ces points sont tels que les frontières $\partial] - \infty, C_h]$ sont disjointes. Nous pouvons alors choisir une suite C_{h_p} de sorte que $L(\partial] - \infty, C_{h_p}] = 0$ pour tout $p \geq 1$ and $C_{h_p} \uparrow (+\infty)^{(k)}$ quand $p \uparrow +\infty$. Ces points sont donc des points de continuité de F et dépassent C . Donc pour tout $p \geq 1$ fixé,

$$F_{n(j)}(C_{h_p}) \geq 1 - \varepsilon.$$

En faisant $j \rightarrow \infty$, nous obtenons

$$F(C_{h_p}) = L([-\infty, C_{h_p}]) \geq 1 - \varepsilon.$$

On conclut en faisant $p \uparrow +\infty$ puis $\varepsilon \downarrow 0$ pour obtenir

$$(3.4) \quad F((+\infty)^{(k)}) = 1.$$

Sur un autre plan, pour tout $\varepsilon > 0$, il existe $M > 0$, tel que

$$\sup P_n(L_M) \leq \varepsilon.$$

Nous devons montrer que

$$(3.5) \quad \lim_{\exists(1 \leq i \leq k), x_i \rightarrow -\infty} F(x) = 0,$$

ce qui revient à dire que pour tout $\varepsilon > 0$, il existe $M > 0$ tel que

$$\exists(1 \leq i \leq k), x_i < -M \implies F(x) \leq \varepsilon.$$

Mais

$$\exists(1 \leq i \leq k), x_i < -M,]-\infty, x] \subset L_M,$$

donc, pour tout $n \geq 1$,

$$\exists(1 \leq i \leq k), x_i < -M, \quad F_n(x) \leq \varepsilon.$$

Maintenant, soit un point x fixé de sorte que : $\exists(1 \leq i \leq k), x_i < -M$. Soit $x(h) = x + h^{(k)}$, avec $0 < h < -(M + x_i)$. Par la méthode, que nous croyons maintenant classique pour nos lecteurs, il existe une suite de points $x(h_p)$, $p \geq 1$, qui sont des points de continuité de F avec $h_p \downarrow 0$. Nous avons donc pour tout $p \geq 1$, fixé pour tout $j \geq 1$,

$$F_{n(j)}(x(h_p)) \leq \varepsilon.$$

En faisant $j \rightarrow \infty$, nous aurons

$$F(x(h_p)) \leq \varepsilon.$$

Maintenant, par continuité à droite, nous avons quand $p \uparrow +\infty$,

$$F(x) \leq \varepsilon.$$

Nous concluons alors : pour tout $\varepsilon > 0$,

$$\exists(1 \leq i \leq k), x_i < -M \implies F(x) \leq \varepsilon.$$

ce qui prouve bien (3.5). Ceci avec (3.4) montre que F est bien une fonction de répartition.

4. Applications

4.1. Théorème de continuité de Lévy.

THÉORÈME 9. *Soit une suite de fonctions caractéristiques ψ_n sur \mathbb{R} convergente vers une fonction ψ continue en zero, alors ψ est une fonction caractéristique.*

PREUVE. Il faut d'abord remarquer que $\psi(0) = 0$ puisque $\psi_n(n)$ pour tout $n \geq 1$. On peut bien supposer les ψ_n sont les fonctions caractéristiques de suites X_n , c'est-à-dire que

$$\psi_n(t) = E(e^{itX_n}), t \in \mathbb{R}.$$

D'après le lemme 11 de la section 2 du chapitre 6, nous avons $|\sin a| \leq a$ pour $|a| \geq 2$. Et donc

$$1_{(|\delta x| > 2)} \leq 2(1 - \frac{\sin \delta x}{\delta x})$$

et, par l'égalité de droite facile à montrer, nous obtenons

$$1_{(|\delta x| > 2)} \leq 2(1 - \frac{\sin \delta x}{\delta x}) = \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos tx) dt.$$

Appliquons cette formule à X_n pour avoir

$$1_{(|X_n| > 2/\delta)} \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos tX_n) dt$$

et en prenant les espérances mathématiques et en appliquant le théorème de Fubini pour des fonctions intégrables, nous arrivons à

$$\mathbb{P}(|X_n| > \frac{2}{\delta}) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} R_e(1 - Ee^{itX_n}) dt.$$

En appliquant le théorème de convergence dominé à $R_e(1 - Ee^{itX_n}) \rightarrow R_e(1 - \psi(t))$, on a

$$(4.1) \quad \liminf_{n \rightarrow \infty} P(|X_n| > \frac{2}{\delta}) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} R_e(1 - \psi(t)) dt.$$

La fonction *partie réelle*, $R_e(\cdot)$ est continue et donc, par hypothèse, $R_e(1 - \psi(t)) \rightarrow 0$ lorsque $\delta \rightarrow 0$. Dès lors lorsque $\delta \rightarrow 0$ dans (4.1), nous obtenons

$$\liminf_{n \rightarrow \infty} P(|X_n| > \frac{2}{\delta}) = 0.$$

Cela implique que la suite est uniformément tendue. Donc elle contient une sous-suite X_{n_k} qui converge vers une variable X . Par le théorème de convergence dominée, pour tout $t \in \mathbb{R}$,

$$\psi_{n_k}(t) = E(\exp(itX_{n_k})) \rightarrow E(\exp(itX)) = \psi_0(t).$$

Par l'unicité des limites dans \mathbb{R} ,

$$\psi = \psi_0$$

et donc ψ est bien une fonction caractéristique.

Passons à la caractérisation de la convergence vague par les fonctions caractéristiques.

4.2. Une autre preuve de la caractérisation de la convergence vague par les fonctions caractéristiques.

THÉORÈME 10. *Une suite variables aléatoires X_n à valeurs dans \mathbb{R}^k converge vaguement vers le vecueur aléatoire $X \in \mathbb{R}^k$ si et seulement si pour tout $u \in \mathbb{R}^k$, $E(\exp(i \langle u, X_n \rangle)) \rightarrow E(\exp(i \langle u, X \rangle))$.*

PREUVE. Le sens direct découle du théorème de convergence dominée. Prouvons le sens indirect et supposons que pour tout $u \in \mathbb{R}^k$

$$\psi_n(u) = E(\exp(i \langle u, X_n \rangle)) \rightarrow E(\exp(i \langle u, X \rangle)) = \psi(u).$$

Pour tout i fixé, $1 \leq i \leq k$, la suite des i -ième composantes $X_n^{(i)}$ verifie, pour tout $t \in \mathbb{R}$,

$$\psi_{n^{(i)}}(t) = \underbrace{\psi_n(0, \dots, t, \dots, 0)}_{i\text{-ième place}} = E(\exp(itX_n^{(i)})) \rightarrow \underbrace{\psi(0, \dots, t, \dots, 0)}_{i\text{-ième place}} = \psi^{(i)}(t).$$

La fonction ψ est continue en zéro puisqu'elle est une fonction caractéristique. Dès lors, les fonctions partielles sont aussi continues en zéro. La formule ci-dessus dit que la suite des fonctions caractéristiques $\psi_{n^{(i)}}(t) = E(\exp(itX_n^{(i)}))$ converge vers une fonction $\psi^{(i)}(t)$ continue en zéro. Le théorème précédent nous fait conclure que la suite $X_n^{(i)}$ est tendue. Ainsi les suites des composantes de X_n sont tendues. Par la proposition 26, la suite X_n est tendue.

Nous pouvons conclure en deux étapes.

Etape 1 : Chaque sous-suite de X_n contient une sous-suite qui converge vaguement vers une certaine mesure de probabilité L . Par hypothèse, X_n converge vers \mathbb{P}_X . Par la caractérisation des lois de probabilités dans et par unicité de la limite vague en loi, nous avons donc

$L = \mathbb{P}_X$. Ainsi, il existe une mesure de probabilité (qui est $L_0 = \mathbb{P}_X$) telle que toute sous-suite X_n contient une sous-suite qui converge vers \mathbb{P}_X .

Etape 2 : Soit $f : \mathbb{R}^k \mapsto \mathbb{R}$ une fonction continue et bornée. Soit une sous-suite $\mathbb{E}f(X_{n_j})$, $j \geq 1$, de la suite $\mathbb{E}f(X_n)$, $n \geq 1$.

Cette sous-suite X_{n_j} , $j \geq 1$, contient une sous-suite $X_{n_{j_\ell}}$, $\ell \geq 1$, qui converge vers L_0 quand $\ell \rightarrow +\infty$. Donc $\mathbb{E}f(X_{n_{j_\ell}})$ converge vers $\int f dL_0$. Ainsi $A = \int f dL_0$ est un nombre réel tel que toute sous-suite $\mathbb{E}f(X_n)$, $n \geq 1$, contient une sous-suite convergente vers A .

Par le critère de Prohorov (Voir Exercice 4 de la section 1 du chapitre 6), $\mathbb{E}f(X_n)$ converge vers $\int f dL_0$. Donc X_n converge vaguement vers $L_0 = P_X$

CHAPTER 4

Outils Particuliers pour la Convergence Vague dans \mathbb{R}

Ce chapitre se focalise sur des outils spécifiques à la convergence vague de suites de variables aléatoires réelles. Pour de telles variables, nous pouvons utiliser les représentations dites de Renyi par le biais de variables aléatoires exponentielles standard ou uniformes standard. De telles représentations sont basées sur les fonctions inverses généralisées sur les quelles le premier chapitre se concentre.

En plus, en relation avec les résultats de la section 6 et du théorème 5 du chapitre 2, le traitement de la convergence vague sur le même espace de probabilité peut devenir une affaire de calculs directs et plus ou moins automatique. Ce chapitre donne des outils dans ce sens.

1. Inverses généralisées des fonctions monotones

Cette théorie est faite pour les fonctions croissantes et continues à droite. Elle peut aussi être faite pour les fonctions décroissantes et continues à gauche. Mais il suffit de la faire pour un des cas de monotonie et d'adapter les définitions et les résultats à l'autre cas.

Soit F be une fonction croissante (au sens large) et continue à droite définie de \mathbb{R} vers \mathbb{R} . Définissons l'inverse généralisée de F comme suit :

$$F^{-1}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}, u \in \mathbb{R}.$$

En raison de l'importance de cette transformation, dite des quantiles, pour la théorie univariée des valeurs extrêmes, nous allons exposer une liste de quelque de ses propriétés importantes. Puisque nous voulons que ces propriétés soient retenues une fois pour toute, nous allons les donner sous formes de points et de pourvoir les preuves à la fin de la liste.

Point (1) Pour tout $u \in \mathbb{R}$ et Pour tout $t \in \mathbb{R}$

$$(A) \quad F(F^{-1}(u)) \geq u$$

et

$$(B) \quad F^{-1}(F(x)) \leq x.$$

Point (2) Pour tout $(u, t) \in \mathbb{R}^2$,

$$(A) \quad (F^{-1}(u) \leq t) \iff (u \leq F(t))$$

et

$$(B) \quad (F^{-1}(u) > t) \iff (u > F(t))$$

Point (3) F^{-1} est croissante et continue à gauche.

Point (4) Définissons la convergence vague d'une suite de fonction monotones $((F_n)_{n \geq 1})$ vers F quand $n \rightarrow +\infty$, noté par $F_n \rightsquigarrow F$, si et seulement si :

$$\forall (x \in C(F)), F_n(x) \rightarrow F(x) \text{ as } n \rightarrow +\infty,$$

où $C(F)$ désigne l'ensemble des points de continuité de F . Nous avons : quand $n \rightarrow +\infty$.

$$(F_n \rightsquigarrow F) \Rightarrow (F_n^{-1} \rightsquigarrow F^{-1})$$

Point (5) Supposons que F_n et F soient **les fonctions de distribution de variables aléatoires réelles** et $F_n \rightsquigarrow F$. Si F est **continue**, alors nous avons la convergence uniforme

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0 \text{ as } n \rightarrow +\infty$$

Point (6) Une fonction de distribution F sur \mathbb{R} a, au plus, un nombre dénombrable de points de discontinuité.

Point (7) Soit \mathbb{P} une probabilité quelconque sur \mathbb{R} avec un support (a, b) , ce qui signifie que $\mathbb{P}((a, b)^c) = 0$ ou ce qui signifie que

$$a = \inf\{x, \mathbb{P}(] - \infty, x]) > 0\} \text{ and } b = \inf\{x, \mathbb{P}(] - \infty, x]) = 1\}.$$

Alors, pour $0 < \varepsilon < 1$, il existe un nombre fini de partition de (a, b) ,

$$a = t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} = b$$

tels que pour $0 < i < k$,

$$\mathbb{P}(]t_i, t_{i+1}[) \leq \varepsilon.$$

Nous pouvons toujours étendre les limites à

$$-\infty \leq t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} \leq +\infty$$

puisque $\mathbb{P}(]-\infty, a[) = 0$ et $\mathbb{P}(]b, +\infty[) = 0$.

Point (8) Soient F et G deux fonctions de distributions à la fois décroissante ou à la fois croissante. Si aucune d'entre elles n'est dégénéré, alors il existe deux points de continuité de toutes les deux F , notés G x_1 et x_2 tels que $x_1 < x_2$ et

$$F(x_1) < F(x_2) \text{ and } G(x_1) < G(x_2).$$

Point (9) Soit F une fonction croissante de \mathbb{R} to $[a, b]$, sans hypothèse de continuité à gauche ou à droite. Alors, pour $y \in]a, b[$,

$$F(F^{-1}(y) - 0) \leq y \leq F(F^{-1}(y) + 0),$$

où $F(\cdot + 0)$ and $F(\cdot - 0)$ désignent respectivement la limite à droite et la limite à gauche de x .

Si la fonction est décroissante, l'inverse généralisée sera définie par

$$F^{-1}(y) = \inf\{x \in \mathbb{R}, F(x) \leq y\}, y \in (a, b).$$

Dans ce cas, nous avons pour tout $y \in (a, b)$

$$F(F^{-1}(y) +) \leq y \leq F(F^{-1}(y) -)$$

B - Preuves des différents points.

Preuves du Point 1. Partie (A). Posons

$$A_u = \{x \in \mathbb{R}, F(x) \geq u\}, u \in \mathbb{R}.$$

Comme $F^{-1}(u) = \inf A_n$, il existe une suite $(x_n)_{n \geq 0} \in A_u$ tels que

$$\begin{cases} F(x_n) & \geq & u \\ x_n & \downarrow & F^{-1}(u) \end{cases}$$

Par la continuité à droite de F , nous avons

$$F(F^{-1}(u)) \geq u.$$

Cela prouve la première formule (A). En ce qui concerne la formule (B), considérons $x \in \mathbb{R}$ et posons

$$F^{-1}(F(x)) = \inf A_{F(x)}$$

En divisant $A_{F(x)}$ dans

$$\begin{aligned} A_{F(x)} &= [-\infty, x[\cap A_{F(x)} + [x, +\infty] \cap A_{F(x)} \\ &=: A_{F(x)}(1) + A_{F(x)}(2) \end{aligned}$$

Par le **fait 1** à la fin de ce paragraphe, nous avons

$$\inf A_{F(x)} = \min(\inf A_{F(x)}(1), \inf A_{F(x)}(2))$$

Mais, nous avons

$$y \in A_{F(x)}(1) \implies y \leq x, \text{ then } \inf A_{F(x)}(1) \leq x$$

Ensuite, nous avons évidemment

$$\inf A_{F(x)}(2) = \{y \geq x, F(y) \geq F(x)\} = x.$$

Ainsi, il vient que

$$\inf A_{F(x)} \leq x.$$

C'est-à-dire:

$$F^{-1}(F(x)) = \inf A_{F(x)} \leq x.$$

Cela met fin à la preuve du point 1 .

Preuves du Point 2. Il est évident que chacune des formules (A) et (B) peut être dérivée l'une de l'autre en prenant les complémentaires. Donc nous pouvons seulement prouver l'une d'elles, par exemple (B). Supposons que $(u > F(t))$. Par continuité à droite de F en t , nous pouvons trouver ϵ tels que

$$u > F(t + \epsilon).$$

Maintenant, pour $x \in A_u$ nous avons sûrement

$$x > t + \epsilon.$$

Sinon, nous aurions

$$x \leq t + \epsilon \implies F(x) \leq F(t + \epsilon) < u,$$

et cela aurait conduit à la conclusion que $x \notin A_u$, qui est en contradiction avec l'hypothèse. Alors, $x > t + \epsilon$ pour tout $x \in A_u$. Cela implique que

$$\inf A_u = F^{-1}(u) \geq t + \epsilon > t.$$

Nous avons prouvé les sens direct de la première formule. Pour prouver les sens indirect, considérons $F^{-1}(u) > t$. Ensuite, supposons que $u > F(t)$ n'est pas vérifié. Cela implique que $F(t) \geq u$, et par suite $t \in A_u$. Puisque

$$\inf A_u = F^{-1}(u) \leq t,$$

ce qui est contraire à l'hypothèse. Donc $u > F(t)$.

Preuves du Point 3. Nous commençons par établir que F^{-1} est non décroissante. Nous avons

$$\forall u \leq u', A_{u'} \leq A_u \implies \inf A_{u'} \leq \inf A_u.$$

Ceci implique que

$$F^{-1}(u') \geq F^{-1}(u)$$

Ensuite, nous devons prouver que F^{-1} est continue à gauche. Soit $u \in \mathbb{R}$. Nous avons pour tout $h \geq 0$,

$$F^{-1}(u - h) \leq F^{-1}(u).$$

Ainsi

$$\lim_{h \downarrow 0} F^{-1}(u - h) \leq F^{-1}(u).$$

Supposons que

$$\lim_{h \downarrow 0} F^{-1}(u - h) = \alpha < F^{-1}(u).$$

Nous pouvons trouver $\epsilon > 0$ tels que $\alpha + \epsilon < F^{-1}(u)$. Maintenant, pour tout $h \geq 0$,

$$F^{-1}(u - h) < \alpha + \epsilon.$$

Par définition de l'infimum, il existe x tel que

$$F(x) \geq u - h \text{ and } F^{-1}(u - h) < \alpha + \epsilon.$$

Par la formule (A) du Point 1,

$$F^{-1}(u - h) < \alpha + \epsilon \implies u - h \leq F(\alpha + \epsilon).$$

Ensuite, nous obtenons comme $h \downarrow 0$

$$u \leq F(\alpha + \epsilon).$$

Etant donné que cela est vrai pour tout $\epsilon > 0$, nous pouvons faire tendre $\epsilon \downarrow 0$ pour obtenir

$$u \leq F(\alpha).$$

Mais, par la formule (B) du Point (2) et en utilisant l'hypothèse, nous arrivons à

$$(\alpha < F^1(u)) \Leftrightarrow (F^{-1}(u) > \alpha) \Leftrightarrow (u > F(\alpha)).$$

Ceci est clairement une contradiction. Nous concluons que

$$\lim_{h \downarrow 0} F^{-1}(u - h) = F^{-1}(u).$$

Et ensuite F^{-1} est continue à gauche.

Preuves du Point 4

Supposons que $F_n \rightsquigarrow F$. Soit $y \in \mathbb{R}$ et soit $\varepsilon > 0$. Comme le nombre de points de discontinuité de F est au plus dénombrable, nous pouvons trouver un point de continuité x de F dans un intervalle ouvert $(F^{-1}(y) - \varepsilon, F^{-1}(y))$. Par le Point 2, $(F^{-1}(y) > x)$ est équivalent à $(F(x) < y)$. Comme $x \in C(F)$, $F_n(x) \rightarrow F(x)$. Alors pour toute valeur de n assez grande, nous avons $F_n(x) < y$ et alors $x < F_n^{-1}(y)$. Nous avons

$$F^{-1}(y) - \varepsilon \leq x < F_n^{-1}(y)$$

c'est-à-dire pour tout $\varepsilon > 0$,

$$F_n^{-1}(y) > F^{-1}(y) - \varepsilon.$$

Faisons n tendre vers $+\infty$ et ε décroître vers 0, nous obtenons pour tout $y \in \mathbb{R}$,

$$\liminf_{n \rightarrow \infty} F_n^{-1}(y) \geq F^{-1}(y).$$

Maintenant, soit y un point de continuité de F^{-1} . Pour tout $y' > y$, nous pouvons trouver un point de continuité point x de F tels que

$$(1.1) \quad F^{-1}(y') < x < F^{-1}(y') + \varepsilon.$$

Par le point 1, $x > F^{-1}(y') \implies F(x) \geq F(F^{-1}(y')) \geq y'$. Alors

$$y < y' \leq F(x).$$

Puisque $x \in C''F$, $F_n(x) \rightarrow F(x)$, nous avons pour les grandes valeurs de n , $y < F_n(x)$ et par la Formule (A) du Point 2, $F_n^{-1}(y) \leq x$. En combinant cela avec (1.1), nous obtenons

$$F^{-1}(y') \geq x \geq F_n^{-1}(y).$$

Maintenant, soit $n \rightarrow +\infty$ pour obtenir

$$\limsup_{n \rightarrow \infty} F_n^{-1}(y) \leq F^{-1}(y').$$

Maintenant, soit $y' \downarrow y$, et nous avons $F^{-1}(y') \downarrow F^{-1}(y)$ par la continuité de F^{-1} en y . Nous arrivons à

$$\limsup_{n \rightarrow \infty} F_n^{-1}(y) \leq F^{-1}(y) \leq \liminf_{n \rightarrow \infty} F_n^{-1}(y).$$

Nous concluons enfin

$$F^{-1}(y) = \limsup_{n \rightarrow \infty} F_n^{-1}(y) = \liminf_{n \rightarrow \infty} F_n^{-1}(y).$$

Preuve du Point 5. Comme F est croissante, x est un point de discontinuité de F si et seulement si le saut de discontinuité $F(x+) - F(x-)$ est strictement positif. Notons par D l'ensemble de tous les points de discontinuité de F , et pour tout $k \geq 1$, notons par D_k l'ensemble des points de discontinuité tels que $F(x+) - F(x-) > 1/k$ et par $D_{k,n}$ l'ensemble des points de discontinuité dans l'intervalle $[-n, n]$ tels que $F(x+) - F(x-) > 1/k$. Nous allons montrer que $D_{k,n}$ est fini.

Supposons que nous pouvons trouver m points x_1, \dots, x_m dans $D_{k,n}$. Comme F est croissante, on peut voir que la somme des sauts de discontinuité est inférieure à $F(n) - F(-n)$. On peut faire un dessin simple pour $m = 3$ et projeter les sauts sur l'axe des ordonnées pour voir cela facilement. Ainsi,

$$\sum_{1 \leq j \leq m} F(x_j+) - F(x_j-) \leq F(n) - F(-n).$$

Puisque chacun de ces sauts dépasse $1/k$, nous avons

$$\begin{aligned} \sum_{1 \leq j \leq m} (1/k) &\leq \sum_{1 \leq j \leq m} F(x_j+) - F(x_j-) \leq F(n) - F(-n). \\ m/k &\leq F(n) - F(-n) \end{aligned}$$

c'est-à-dire

$$m \leq k(F(n) - F(-n))$$

Nous concluons en disant que nous ne pouvons pas avoir plus que $[k(F(n) - F(-n))]$ points dans $D_{k,n}$, donc $D_{k,n}$ est fini. Comme

$$D = \cup_{n \geq 1} \cup_{k \geq 1} D(k, n)$$

Nous voyons que D est dénombrable. Ceci met fin à la preuve.

Preuve du Point 6. Soit $0 < \varepsilon < 1$. Let $F(t) = \mathbb{P}(] - \infty, t])$. Ceci est une fonction de répartition, donc vérifiant $F(\infty) = 0$ et $F(+\infty) = 1$. Fixons $0 < \varepsilon < 1$. Posons $k = [1/\varepsilon]$, où $[t]$ représente la partie entière de t . Nous avons alors

$$k\varepsilon \leq 1 \leq k\varepsilon + \varepsilon.$$

Notons

$$s_i = i\varepsilon, \text{ for } i=1, \dots, k \text{ and } s_{k+1} = 1.$$

et

$$t_i = F^{-1}(s_i) = \inf\{u, G(u) \geq s_i\}.$$

Par le Point 1,

$$(1.2) \quad F(t_i) \geq s_i.$$

Ensuite, pour tout $1 \leq i < k$,

$$F(t_{i+1}-) = \lim_{h \downarrow 0} F(t_{i+1} - h).$$

Par définition de t_{i+1} , qui est l'infimum des valeurs de u tels que $F(u) \geq (i+1)\varepsilon$, nous avons sûrement,

$$F(t_{i+1} - h) < (i+1)\varepsilon.$$

En allant à la limite, nous obtenons

$$(1.3) \quad B(t_{i+1}-) \leq (i+1)\varepsilon.$$

En mettant ensemble (1.2) et (1.3), nous avons

$$\mathbb{P}(]t_i, t_{i+1}[) = F(t_{i+1}-) - F(t_i) \leq (i+1)\varepsilon - i\varepsilon = \varepsilon,$$

pour $i = 1, \dots, k$. Enfin $i = k$, nous avons $F(t_{k+1}) = 1$ et

$$\mathbb{P}(]t_k, t_{k+1}[) = 1 - F(t_k) \leq 1 - k\varepsilon \leq \varepsilon.$$

Pour $i = 0$, comme $F(t_0) \geq 0$, nous avons

$$\mathbb{P}(]t_0, t_1[) = F(t_1-) - F(t_0) \leq F(t_1-) \leq \varepsilon$$

Nous venons de prouver que $0 \leq i \leq k$,

$$\mathbb{P}(]t_i, t_{i+1}[) = F(t_{i+1}+) - F(t_i) \leq (i+1)\varepsilon - i\varepsilon = \varepsilon.$$

Preuve du Point 7. Nous allons appliquer le Point 6. Considérons la mesure de probabilité de Lebesgues-Stieljes générée par F et caractérisée par

$$\mathbb{P}(]u, v]) = F(v) - F(u), u \leq v.$$

En particulier, nous avons $\mathbb{P}(]-\infty, v]) = F(v) - F(u)$. Fixons $\varepsilon > 0$ et considérons une subdivision

$$-\infty = t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} = +\infty$$

tel que pour tout $0 \leq j \leq k$,

$$F(t_{j+1}-) - F(t_j) = \mathbb{P}_X(]t_j, t_{j+1}[) \leq \varepsilon.$$

Maintenant, nous voulons prouver la convergence uniforme. Soit x un des t_j . Nous avons

$$F_n(x) - F(x) \leq \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)|.$$

Tout autre x est dans l'un des intervalles $]t_j, t_{j+1}[$. Utilisons la croissance de F et F_n pour avoir

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(t_{j+1}-) - F(x) \\ &\leq F_n(t_{j+1}-) - F(t_{j+1}-) + F(t_{j+1}-) - F(x) \\ &\leq F_n(t_{j+1}-) - F(t_{j+1}-) + F(t_{j+1}-) - F(t_j) \\ &\leq \sup_{0 \leq j \leq k+1} |F_n(t_j-) - F(t_j-)| + \varepsilon \end{aligned}$$

et

$$\begin{aligned} F(x) - F_n(x) &\leq F(x) - F_n(t_j) \\ &\leq F(x) - F(t_j) + F(t_j) - F_n(t_j) \\ &\leq F(t_{j+1}-) - F(t_j) + F(t_j) - F_n(t_j) \\ &\leq \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)| + \varepsilon \end{aligned}$$

A l'arrivée, nous avons pour tout x différent des t_j ,

$$|F(x) - F_n(x)| \leq \max\left(\sup_{0 \leq j \leq k+1} |F_n(t_j-) - F(t_j-)|, \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)|\right) + \varepsilon.$$

Alors, en notant,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \|F_n - F\|_\infty,$$

nous avons obtenu

$$\|F_n - F\|_\infty \leq \max\left(\sup_{0 \leq j \leq k+1} |F_n(t_j-) - F(t_j-)|, \sup_{0 \leq j \leq k+1} |F_n(t_j) - F(t_j)|\right) + \varepsilon.$$

A ce stade, nous avons la conclusion plus générale. Si pour tous les réels x , $F_n(x) \rightarrow F(x)$ et $F_n(x-) \rightarrow F(x-)$, alors nous pouvons conclure que

$$\limsup \|F_n - F\|_\infty \leq \varepsilon,$$

pour un $\varepsilon > 0$ arbitraire. Ainsi, nous avons

$$\lim \|F_n - F\|_\infty = 0.$$

Pour étendre cette conclusion pour le cas où F est continue et $F_n(x) \rightarrow F(x)$, nous allons prouver que $F_n(x-) \rightarrow F(x)$ pour tout x .

Pour le prouver cela pour un quelconque x fixé et avec $0 \leq h_p \downarrow 0$ comme $p \uparrow +\infty$. Nous avons pour chaque n ,

$$F_n(x-) - F(x) \leq F_n(x) - F(x) \leq |F_n(x-) - F(x)|$$

et

$$F(x) - F_n(x-) \leq F(x) - F_n(x - h_p) \leq |F(x) - F(x - h_p)| + |F(x - h_p) - F_n(x - h_p)|$$

En combinant ces deux points, nous avons

$$|F(x) - F_n(x-)| \leq \max(|F_n(x-) - F(x)|, |F(x) - F(x - h_p)| + |F(x - h_p) - F_n(x - h_p)|).$$

Maintenant fixons p et soit $n \rightarrow +\infty$ pour avoir

$$\limsup_{n \rightarrow +\infty} |F(x) - F_n(x-)| \leq |F(x) - F(x - h_p)|.$$

Finalement, faisons $p \rightarrow +\infty$ pour avoir la conclusion par le fait de la continuité de F .

Preuve du Point 8. F est dégénérée si et seulement si il y a un unique point de croissance, par exemple a , en lequel elle présente un saut de discontinuité. Il est sous la forme : $F(x) = c_1$ pour $x < a$ et

$F(x) = c_2$ pour $x \geq a$, avec $c_1 < c_2$. Donc, si F est non dégénérée, il comporte au moins deux points de croissance. Par conséquent, nous pouvons trouver trois points de continuité de F : $a_1 < a_1 < a_3$ tels que $F(a_1) < F(a_1) < F(a_3)$. Si G est également non dégénérée, on peut trouver aussi trois points de continuité de G : $b_1 < b_1 < b_3$ tels que $F(b_1) < F(b_1) < F(b_3)$. Nous considérons deux cas.

Cas 1. L'intervalle $[a_1, a_3]$ et $[b_1, b_3]$ sont disjoints ou, ont une intersection réduite à un point qui forcément un point de continuité. Supposons par exemple que $a_3 \leq b_1$. Prenons $x_1 = a_1$ et $x_2 = b_3$. Nous avons

$$F(x_1) < F(a_3) \leq F(b_1) \leq F(b_3) = F(x_2)$$

et

$$G(x_1) \leq G(a_3) \leq G(b_1) < F(b_3) = G(x_2).$$

Cas 2. L'intervalle $[a_1, a_3]$ et $[b_1, b_3]$ ont une intersection non vide et non réduite à un point. Cette intersection possède un intérieur non vide. On peut donc y choisir un point t de continuité de F et de G . Nous avons alors sûrement $F(a_1) < F(t)$ ou $F(t) < F(a_3)$. Sinon, nous aurions $F(a_1) = F(a_3)$, ce qui violerait ce qui est ci-dessus. De façon similaire $G(b_1) < G(t)$ ou $G(t) < G(a_3)$.

Maintenant, prenons $x_1 = \min(a_1, b_1)$ et $x_2 = \min(a_3, b_3)$. Discutons :

Si $F(a_1) < F(t)$, nous avons

$$F(x_1) \leq F(a_1) < F(t) \leq F(a_3) \leq F(x_2)$$

Si $F(t) < F(a_3)$, nous avons

$$F(x_1) \leq F(a_1) \leq F(t) < F(a_3) \leq F(x_2)$$

Nous concluons que

$$F(x_1) < F(x_2)$$

Nous montrons de façon similaire que

$$G(x_1) < G(x_2)$$

Par le Point 5, nous savons que les points de discontinuité de F et G sont au plus dénombrable. Nous, nous pouvons ajuster x_1 et x_2 de

sorte qu'ils soient à la fois des points de continuité de F et de G .

Preuve du Point (9). Commençons par le premier dans lequel la fonction F est croissante. Par définition de l'inverse généralisée, nous avons pour tout $h > 0$

$$F(F^{-1}(y) + h) \geq y$$

et and

$$F(F^{-1}(y) - h) < y.$$

En passant à la limite quand $h \downarrow 0$, nous obtenons

$$F(F^{-1}(y)-) \leq y \leq F(F^{-1}(y)+).$$

Similairement, si F est décroissante, nous avons pour tout $h > 0$,

$$F(F^{-1}(y) + h) \leq y$$

et

$$F(F^{-1}(y) - h) > y.$$

En passant à la limite quand $h \downarrow 0$, nous obtenons

$$F(F^{-1}(y)+) \leq y \leq F(F^{-1}(y)-), \quad y \in (a, b).$$

Fact 1. Soit A et B deux sous-ensembles disjoint de \mathbb{R} . Nous avons have

$$\inf A \cup B = \min(\inf A, \inf B).$$

En effet, de manière évidente, $\inf A \cup B$ est inférieur à $\inf A$ et à $\inf B$, et par suite, nous avons $\inf A \cup B \leq \min(\inf A, \inf B)$. A l'inverse, supposons ue nous n'avons pas l'égalité, i.e.,

$$\inf A \cup B < \min(\inf A, \inf B).$$

Dans ce cas, il existerait une suite $(z_n)_{n \leq 0}$ de points $A \cup B$ décroissante vers $\inf(A \cup B)$. Alorsn sûrement, lorsque n est assez grand, z_n sera strictement plus petit que chacun des nombres $\inf A$ et $\inf B$. Et pourtant, il soit soit dans A soit dans B . Ceci est absurde. Nous concluons que nous avons l'égalité.

1.1. Applications des fonctions inverses généralisées. La première application est la représentation d'une variable aléatoire réelle par une variable aléatoire uniforme standard $U \sim \mathcal{U}(0, 1)$ associée à la fonction de répartition $H(x) = 0$ pour $x < 0$, $H(x) = x$ pour $x \in (0, 1)$ et $H(x) = 1$ pour $x > 0$. Nous avons :

LEMME 4. *Soit F une fonction de distribution tels que $F(-\infty) = 0$ et $F(+\infty) = 1$. Soit $U \sim \mathcal{U}(0, 1)$, définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Alors F est la fonction de répartition de $X = F^{-1}(U)$.*

Preuve. Nous avons par la formule (A) du point 2 ci-dessus que

$$\mathbb{P}(X \leq x) = P(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Une deuxième application est cette forme simple du Théorème de Skorohod.

THÉORÈME 11. *Soit $F_n \rightsquigarrow F$, où F_n et F sont des fonction de distribution tels que $F_n(-\infty) = 0$ et $F_n(+\infty) = 1$, pour $n \geq 0$, $F(-\infty) = 0$ et $F(+\infty) = 1$. Alors, il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ portant une suite variables aléatoires réelles X_n et une variable aléatoire X telles que pour tout $n \geq 0$, F_n est la fonction de répartition de X_n , c'est-à-dire $F_n(\cdot) = \mathbb{P}(X_n \leq \cdot)$, et F est la répartition X , i.e. $F(\cdot) = \mathbb{P}(X \leq \cdot)$ et*

$$X_n \rightarrow p.s \text{ as } n \rightarrow +\infty.$$

Preuve. Considérons $([0, 1], \mathcal{B}([0, 1]), \lambda)$ où λ est la mesure de Lebesgues sur $[0, 1]$, qui est une mesure de probabilité. Considérons la fonction identité : $U : ([0, 1], \mathcal{B}([0, 1]), \lambda) \mapsto ([0, 1], \mathcal{B}([0, 1]), \lambda)$. Alors U suit une loi uniforme standard puisque pour tout $x \in (0, 1)$,

$$\lambda(U \leq x) = \lambda(U^{-1}([-\infty, x]))$$

où U^{-1} est l'inverse de U et alors $U^{-1}([-\infty, x]) =]-\infty, x]$. Ainsi

$$\lambda(U \leq x) = \lambda(U^{-1}([-\infty, x])) = \lambda([-\infty, x]) = x.$$

Considerons $X_n = F_n^{-1}(U)$ et $X = F^{-1}(U)$. Alors chaque F_n est une fonction de répartition de X_n et F est une fonction de répartition de X . Montrons que X_n converge vers X presque sûrement. En utilisant le point 4 ci-dessus, nous avons $F_n^{-1} \rightsquigarrow F^{-1}$.

Nous avons

$$\begin{aligned}
 1 &\geq \lambda(X_n \rightarrow X) = \lambda(\{u \in [0, 1], X_n(u) \rightarrow X(u)\}) \\
 &= \lambda(\{u \in [0, 1], F_n^{-1}(u) \rightarrow F^{-1}(u)\}) \\
 &\geq \lambda(\{u \in [0, 1], u \text{ est un point de continuité de } F\}) = 1,
 \end{aligned}$$

étant donné que le complémentaire de $\{u \in [0, 1], u \text{ est un point de continuité de } F\}$ est dénombrable et la mesure d'un ensemble par la mesure de Lebesgue est nulle.

2. Représentation uniform et exponentielles de Renyi

Dans cette section, nous nous intéressons aux représentations de statistique d'ordre $X_{1,n} \leq \dots \leq X_{n,n}$, $n \geq 1$, de n variables aléatoires indépendantes X_1, \dots, X_n de fonction de répartition commune F par rapport à des variables uniformes ou exponentielles standard.

Dans toute la section, les variables sont définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Nous commençons par rappeler la densité de probabilité (dp) de la statistique d'ordre issue d'un échantillon de dp h .

2.1. Densité de probabilité de la statistique d'ordre.

PROPOSITION 28. *Soit Z_1, Z_2, \dots, Z_n n copies indépendantes d'une variable aléatoire Z de loi absolument continue et de dp h . Alors la statistique d'ordre associée $(Z_{1,n}, Z_{2,n}, \dots, Z_{n,n})$ est de loi absolument continue et sa dp jointe est donnée par*

$$h_{(Z_{1,n}, \dots, Z_{n,n})}(z_1, \dots, z_n) = n! \prod_{i=1}^n h(z_i) 1_{(z_1 \leq \dots \leq z_n)}.$$

Preuve.

Supposons que les hypothèses de la proposition soient vraies. Trouvons la densité conjointe des r statistiques d'ordre $Z_{n_1,n} \leq Z_{n_2,n} \leq \dots \leq Z_{n_r,n}$, avec $1 \leq r \leq n$, $1 \leq n_1 < n_2 < \dots < n_r$.

Puisque la variable aléatoire Z est absolument continue, ses observations sont distinctes presque sûrement et nous avons $Z_{n_1,n} < Z_{n_2,n} < \dots$.

$\dots < Z_{n_r, n}$, *p.s.* Soit dz_i , $1 \leq i \leq r$, assez petits. Alors pour $z_1 < z_2 < \dots < z_r$, l'évènement

$$(2.1) \quad (Z_{n_i, n} \in]z_i - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r).$$

se réalise avec $n_1 - 1$ observations de l'échantillon Z_1, \dots, Z_n tombant strictement à gauche de z_1 , une dans $]z_1 - dz_1/2, z_1 + dz_1/2[$, $n_2 - n_1 - 1$ entre $z_1 + dz_1/2$ et $z_2 - dz_1/2$, une dans $]z_2 - dz_2/2, z_2 + dz_2/2[$, etc., et $n - k_k$ observations strictement à droite de z_r .

Cette description est représentée dans la figure 1 pour $r = 3$.

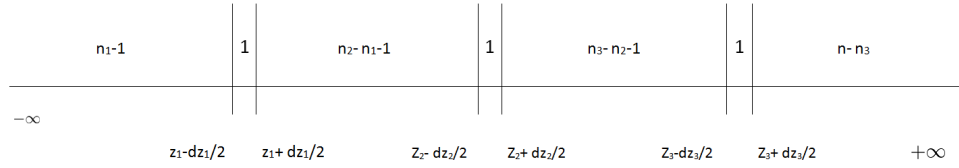


FIGURE 1. Placement des observations aux points $z_1 < \dots < z_r$ lors de l'évènement (2.1) a lieu

Par définition, la densité de probabilité $f_{(Z_{n_1, n}, \dots, Z_{n_r, n})}$, si elle existe, doit vérifier

$$(2.2) \quad \frac{\mathbb{P}(Z_{n_i, n} \in]z_i - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)}{dz_1 \times \dots \times dz_r} = f_{(Z_{n_1, n}, \dots, Z_{n_r, n})}(z_1, \dots, z_r)(1 + \varepsilon(dz_1, \dots, dz_r)),$$

où $\varepsilon(dz_1, \dots, dz_r) \rightarrow 0$ quand si chaque $\Delta_i \rightarrow 0$ ($1 \leq i \leq r$). Maintenant, en utilisant l'idépendance de l'échantillon Z_1, \dots, Z_n , la probabilité $\mathbb{P}(Z_{n_i, n} \in]z_i - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)$ est obtenue avec par une distribution multinomiale. En utilisant enfin le fait que h est la densité de probabilité conjointe des observations, nous obtenons

$$\begin{aligned}
& \frac{\mathbb{P}(Z_{n_i,n} \in]z_1 - dz_i/2, z_i + dz_i/2[, \quad 1 \leq i \leq r)}{dz_1 \times \dots \times dz_r} \\
&= n! \times \frac{h(z_1)^{n_1-1}}{(n_1-1)!} \times \frac{(F(z_2) - F(z_1))^{n_2-n_1-1}}{(n_2-n_1-1)!} \\
&\times \dots \times \frac{(F(z_j) - F(z_{j-1}))^{n_j-n_{j-1}-1}}{((n_j-n_{j-1}-1)!) } \\
&\times \dots \times \frac{(F(z_r) - F(z_{r-1}))^{n_r-n_{r-1}-1}}{(n_r-n_{r-1}-1)!} \\
&\times \frac{(1 - F(z_r))^{n-n_r}}{(n-n_r)!} \\
&\times \prod \frac{\mathbb{P}(Z_{n_i,n} \in]z_i - dz_i/2, z_i + dz_i/2[)}{1! \Delta_i}.
\end{aligned}$$

Le dernier facteur du produit est donné par

$$\prod_{i=1}^r h(z_i)(1 + dz_i).$$

En posant $n_0 = 0$ et $n_r = n + 1$ et pour tout $-\infty = z_0 < z_1 < \dots < z_r < z_{r+1} = +\infty$,

$$f_{(Z_{n_1,n}, \dots, Z_{n_r,n})}(z_1, \dots, z_r) = n! \prod_{j=1}^{r+1} \frac{h(z_j)(F(z_j) - F(z_{j-1}))^{n_j-n_{j-1}-1}}{(n_j-n_{j-1}-1)!},$$

nous voyons que $f_{(Z_{n_1,n}, \dots, Z_{n_r,n})}$ satisfait (2.2). Dès lors, nous avons la conclusion partielle :

LEMME 5. *Supposons que Z_1, Z_2, \dots, Z_n soient n observations indépendantes d'une variable aléatoire Z de loi absolument continue par rapport à la mesure de Lebesgues et densité de probabilité h par rapport à la mesure de Lebesgues et de fonction de répartition H , définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Soit $1 \leq r \leq n$, $1 \leq n_1 < n_2 < \dots < n_r$. Alors les r statistiques d'ordre $Z_{n_1,n} < Z_{n_2,n} < \dots < Z_{n_r,n}$ admettent pour densité de probabilité conjointe par rapport à la mesure de Lebesgues, la fonction ci-dessous définie en (z_1, \dots, z_r) par*

$$(2.3) \quad n! \prod_{j=1}^{r+1} \frac{h(z_j)(F(z_j) - F(z_{j-1}))^{n_j-n_{j-1}-1}}{(n_j-n_{j-1}-1)!} 1_{(z_1 < \dots < z_r)},$$

où, par convention, $n_0 = 0$ et $n_r = n + 1$, $z_0 = -\infty = z_0$ et $z_{r+1} = +\infty$.

Pour finir la preuve, posons $r = n$ et $n_1 = 1, n_2 = 2, \dots, n_n = n$. Puisque les nombres $n_j - n_{j-1} - 1$ s'annulent dans (2.3), il vient que les statistiques $Z_{1,n} < Z_{2,n} < \dots < Z_{n,n}$ ont la densité de probabilité conjointe égale à

$$n! \prod_{j=1}^n h(z_j) \mathbf{1}_{(z_1 < \dots < z_r)},$$

Dans le reste de la section, nous nous focalisons sur les relations entre statistiques d'ordre issues de loi uniforme ou exponentielles standard.

PROPOSITION 29. *Soit $n \geq 1$ fixé et $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$ la statistique d'ordre associée à U_1, U_2, \dots, U_n , n variables aléatoires indépendantes uniformément distribuées sur $(0, 1)$. Soit $E_1, E_2, \dots, E_n, E_{n+1}$, $(n+1)$ une suite de $n+1$ variables aléatoires indépendantes suivant une loi exponentielle standard, i.e.,*

$$\forall x \in \mathbb{R}, \quad \mathbb{P}(E_i \leq x) = (1 - e^{-x}) \mathbf{1}_{(x \geq 0)}, i = 1, \dots, n+1.$$

Soit $S_j = E_1 + \dots + E_j$, $1 \leq j \leq n+1$. Alors nous avons l'égalité en loi suivante :

$$(U_{1;n}, U_{2;n}, \dots, U_{n;n}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right)$$

Preuve. D'une part, d'après (28), la dp conjointe de $U = (U_{1,n}, U_{2,n}, U_{n,n})$ est donnée par

$$\forall (u_1, \dots, u_n) \in \mathbb{R}^n, f_U(u_1, \dots, u_n) = n! \mathbf{1}_{(0 \leq u_1 \leq \dots \leq u_n \leq 1)}.$$

Nous allons trouver la distribution de $Z_{n+1}^* = (S_1, S_2, \dots, S_n, S_{n+1})$ conditionnellement à $S_{n+1} = t$, $t > 0$. Nous avons pour $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$,

$$\begin{aligned} (2.4) \quad f_{Z_n^*}^{S_{n+1}=t}(y) &= \frac{f_{(Z_n^*, S_{n+1})}(y, t)}{f_{S_{n+1}}(t)} \\ &= \frac{f_{Z_{n+1}^*}(y, t)}{f_{S_{n+1}}(t)} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq t)}. \end{aligned}$$

Mais S_{n+1} suit une loi gamma de paramètres $n+1$ et 1, i.e., $S_{n+1} \sim \gamma(n+1, 1)$, et sa dp est :

$$(2.5) \quad f_{S_{n+1}}(t) = \frac{t^n e^{-t}}{\Gamma(n+1)} \mathbf{1}_{(t \geq 0)} = \frac{t^n}{n!} e^{-t} \mathbf{1}_{(t \geq 0)}.$$

Trouvons maintenant la distribution de (S_1, \dots, S_{n+1}) . Elle vient de celle de la transformation

$$\begin{pmatrix} E_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ E_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & & & & & \\ -1 & 1 & & & & \\ 0 & -1 & 1 & & & \\ \cdot & & \cdot & \cdot & & \\ \cdot & & & \cdot & \cdot & \\ \cdot & & & & -1 & 1 \end{pmatrix} \begin{pmatrix} S_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ S_{n+1} \end{pmatrix}$$

Soit B la matrice de la formule ci-dessus. Le jacobien de cette transformation est : $|B| = 1$ and

$$B \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_{n+1} \end{pmatrix} = (y_1, y_2 - y_1, \dots, y_{n+1} - y_n)$$

Donc la dp de (S_1, \dots, S_{n+1}) est la suivante

$$\begin{aligned} f_{Z_{n+1}^n}(y_1, \dots, y_{n+1}) &= f_{(E_1, \dots, E_{n+1})}(B(y_1, \dots, y_{n+1})) \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq y_{n+1})} \\ &= \prod_{i=1}^{n+1} e^{-(y_i - y_{i-1})} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq y_{n+1})} \\ &= e^{-y_{n+1}} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq y_{n+1})}. \end{aligned}$$

avec $y_0 = 0$ par convention. Retournant à (2.4) et à (2.5), nous obtenons avec $y = (y_1, y_2, \dots, y_n)$,

$$(2.6) \quad f_{Z_n^*}^{S_{n+1}=t}(y) = \frac{n!}{t^n} \mathbf{1}_{(0 \leq y_1 \leq \dots \leq y_n \leq t)}.$$

Posons $u = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$, nous avons

$$f_{\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)}^{S_{n+1}=t}(u) = f_{\left(\frac{S_1}{t}, \dots, \frac{S_n}{t}\right)}^{S_{n+1}=t}(u_1, u_2, \dots, u_n).$$

Ceci est la dp obtenue en (2.6) par la transformation

$$(y_1, y_2, \dots, y_n) = t(u_1, u_2, \dots, u_n) \iff (u_1, u_2, \dots, u_n) = \frac{1}{t}(y_1, y_2, \dots, y_n)$$

de jacobien t^n . Dès lors,

$$\begin{aligned} f_{\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}\right)}^{S_{n+1}=t}(u_1, u_2, \dots, u_n) &= f_{Z_n^*}^{S_{n+1}=t}(t(u_1, u_2, \dots, u_n)) t^n \mathbf{1}_{(0 \leq tu_1 \leq \dots \leq tu_n \leq t)} \\ &= n! \mathbf{1}_{(0 \leq u_1 \leq \dots \leq u_n \leq 1)}. \end{aligned}$$

Ceci est exactement (2.4). Nous concluons que la distribution de $Z = \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right)$ conditionnellement à $S_{n+1} = t$ ne dépend pas de t . Il s'en suit que cette distribution conditionnelle est la distribution inconditionnelle si bien que

$$\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right)$$

possède la même loi que $U = (U_{1,n}, U_{2,n}, \dots, U_{n,n})$ et de plus, elle est indépendante S_{n+1} . Ceci finit la preuve.

Nous formalisons la dernière conclusion par

LEMME 6. *Soit $E_1, E_2, \dots, E_n, E_{n+1}, n \geq 1$ n variables aléatoires indépendantes de loi exponentielle standard définies sur le même espace de probabilité. Soit $S_i = E_1 + E + \dots + E_i, 1 \leq i \leq n+1$. Alors*

$$\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right)$$

est indépendant de S_{n+1} .

La proposition précédente exprime une représentation de la statistique d'ordre issue d'une loi uniforme standard en fonction de celles usées d'une loi exponentielle standard. La proposition qui suit propose une voie inverse.

PROPOSITION 30. *Adoptons les notations de la proposition 29. Alors pour tout $n \geq 1$, nous avons*

$$(-\log U_{1,n}, \dots, -\log U_{n,n}) =^d (E_{1,n}, \dots, E_{n,n}),$$

où $E_{1,n} \leq \dots \leq E_{n,n}$ est la statistique d'ordre associée à E_1, E_2, \dots, E_n, n variables aléatoires indépendantes suivant la loi exponentielle standard.

Preuve. Par la proposition 28, la dp de $E_{1,n} \leq \dots \leq E_{n,n}$ est

$$(2.7) \quad f_Z(z) = n! e^{-\sum_{i=1}^n z_i} \mathbf{1}_{(0 \leq z_1 \leq \dots \leq z_n)}, \quad z = (z_1, \dots, z_n) \in \mathbb{R}^n.$$

où $Z = (E_{1,n}, \dots, E_{n,n})$. La distribution de $Z^* = (-\log U_{1,n}, \dots, -\log U_{n,n})$ résulte de celle de $U = (U_{1,n}, \dots, U_{n,n})$ par la transformation difféomorphe $(z_1, \dots, z_n) = (-\log u_1, \dots, -\log u_n)$ qui préserve l'ordre des arguments. Le jacobien de cette transformation est donné par :

$$\begin{aligned} \left| \frac{\partial U_i}{\partial z_j} \right| &= \left| \partial \frac{\partial e^{-z_i}}{\partial z_j} \right| = \left| \text{diag}(-e^{-z_1}, \dots, -e^{-z_n}) \right| \\ &= e^{-\sum_{i=1}^n z_i}. \end{aligned}$$

Dès lors, la pd de Z^* est

$$\begin{aligned} f_{Z^*}(z_1, \dots, z_n) &= f_U(-e^{-z_1}, \dots, -e^{-z_n}) e^{-\sum_{i=1}^n z_i} \mathbf{1}_{(0 \leq z_1 \leq \dots \leq z_n)} \\ &= n! e^{-\sum_{i=1}^n z_i} \mathbf{1}_{(0 \leq z_1 \leq \dots \leq z_n)}. \end{aligned}$$

Cette dpr est exactement celle de $(E_{1,n}, \dots, E_{n,n})$ en vertu de (2.7). Ceci finit la preuve.

Donnons une autre version de ce résultat. Il est évident que pour toute variable aléatoire U suivant la loi uniforme standard, nous avons $U \stackrel{d}{=} 1 - U$. Dès lors, pour tout $n \geq 1$, nous avons

$$(U_{1,n}, \dots, U_{n,n}) \stackrel{d}{=} (1 - U_{1,n}, \dots, 1 - U_{n,n}).$$

L'égalité en loi dans la proposition 30 devient : pour tout $n \geq 1$,

$$(-\log(1 - U_{n,n}), \dots, -\log(1 - U_{1,n})) \stackrel{d}{=} (E_{1,n}, \dots, E_{n,n})$$

Allons plus loin et posons

$$\alpha_{i,n} = -\log(1 - U_{i,n}), 1 \leq i \leq n.$$

Pour tout $n \geq 1$, soit la transformation

$$\begin{pmatrix} n\alpha_{1,n} \\ (n-1)(\alpha_{2,n} - \alpha_{1,n}) \\ \vdots \\ (n-i+1)(\alpha_{i,n} - \alpha_{i-1,n}) \\ \vdots \\ 1(\alpha_{n,n} - \alpha_{n-1,n}) \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_i \\ \vdots \\ V_n \end{pmatrix}.$$

We have

$$\begin{pmatrix} \alpha_{1,n} \\ \alpha_{2,n} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{n,n} \end{pmatrix} = \begin{pmatrix} V_1/n \\ V_1/n + V_2/(n-1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ V_1/n + V_2/(n-1) + \dots + V_{n-1}/2 + V_1/1 \end{pmatrix}$$

Par la formule de changement de variable classique, la dp de (V_1, \dots, V_n) est donnée par

$$f_V(v_1, \dots, v_n) = f_{(\alpha_{1,n}, \dots, \alpha_{n,n})}(v_1/n, v_1/n + v_2/(n-1), \dots, v_1/n + v_2/(n-1) + \dots + v_n) \times |J(v)| \times \mathbf{1}_{D_V}(v).$$

La jacobien de cette transformation est donnée par

$$|J(v)| = \frac{1}{n!}$$

et le support de V est

$$D_V = \mathbb{R}_+^n$$

Nous pouvons conclure en utilisant (2.7) qui donne la dp de $(\alpha_{1,n}, \dots, \alpha_{n,n})$, et en posant $s_i = v_1/n + v_2/(n-1) + \dots + v_i/(n-i+1)$, $i = 1, \dots, n$. Nous obtenons

$$\begin{aligned} f_V(v_1, \dots, v_n) &= \frac{1}{n!} \times n! e^{-\sum_{i=1}^n s_i} \mathbf{1}_{(v_1 \geq 0, \dots, v_n \geq 0)} \\ &= e^{-\sum_{i=1}^n s_i} \mathbf{1}_{(v_1 \geq 0, \dots, v_n \geq 0)} \end{aligned}$$

Nous pouvons vérifier facilement que $s_1 + \dots + s_n = v_1 + \dots + v_n$. Tout ceci nous amène à

$$f_V(v_1, \dots, v_n) = \prod_{i=1}^n e^{-v_i} \mathbf{1}_{(v_i \geq 0)}.$$

Ceci nous dit que le vecteur (V_1, \dots, V_n) est à composantes indépendantes suivant chacune une loi exponentielle standard. Résumons ce fait par :

PROPOSITION 31. *Adoptons les notations précédentes. Soit $\alpha_{i,n} = -\log(1 - U_{i,n})$, $i = 1, \dots, n$, pour $n \geq 1$. Alors, les variables aléatoires $n\alpha_{1,n}, (n-1)(\alpha_{2,n} - \alpha_{1,n}), \dots, (n-i+1)(\alpha_{i,n} - \alpha_{i-1,n}), \dots, (\alpha_{n,n} - \alpha_{n-1,n})$ sont indépendantes et suivent la loi exponentielle standard.*

Allons plus loin et posons $1 \leq i \leq n$,

$$(n - i + 1) (\alpha_{i,n} - \alpha_{i-1,n}) = (n - i + 1) \log \left(\frac{1 - U_{i-1,n}}{1 - U_{i,n}} \right).$$

De ce qui précède, nous avons que les variables aléatoires

$$E_{n-i+1}^* = (n - i + 1) (\alpha_{i,n} - \alpha_{i-1,n}) = \log \left(\frac{1 - U_{n-i,n}}{1 - U_{n-i+1,n}} \right)^{(n-i+1)}$$

sont indépendantes et suivent la loi exponentielle standard. Utilisons la représentation distributionnelle des $U_{n-i,n}$ par les $U_{i+1,n}$ pour arriver à :

PROPOSITION 32. (*Malmquist representation*). Soit U_1, U_2, \dots, U_n une suite de $n \geq 1$ variables aléatoires indépendantes et suivant toute une loi uniforme standard. Soit $0 \leq U_{1,n} < U_{2,n} < \dots < U_{n,n} \leq 1$ la statistique d'ordre associée. Alors, les variables aléatoires

$$\log \left(\frac{U_{i+1,n}}{U_{i,n}} \right)^i, i = 1, \dots, n$$

sont indépendantes et suivent la loi exponentielle standard.

Le processus empirique fonctionnel comme outil général en statistique asymptotique

1. Utilisation du petit o et du grand O

Dans ce chapitre, nous montrerons comment ombiner tous les concepts que nous avons étudié aussi loin pour obtenir des outils encore simples mais puissants qui peuvent être systématiquement utilisés pour trouver des lois asymptotiques normales dans une grande variété de problèmes, même dans de problèmes de recherche d'au jour d'aujourd'hui. Nous étudierons d'abord les manipulations des symboles $o_{\mathbb{P}}$ et $O_{\mathbb{P}}$ concernant les limites en probabilité. Ensuite, nous présentons le processus empirique fonctionnel qui est utilisé ici seulement dans le contexte du cas des distributions finies. Et, nous donnerons quelques cas comme illustrations.

Il est important de noter pour une fois que la méthode donnée est valide pour les suites de variables aléatoires et la limite de variables aléatoires définie sur le même espace de probabilité. En conséquence, nous traitons des suites de variables aléatoires $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$, $(Z_n)_{n \geq 1}, \dots$ définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^k , $k \geq 1$ et $(a_n)_{n \geq 1}$, $(b_n)_{n \geq 1}$, $(c_n)_{n \geq 1}$ sont des nombres aléatoires strictement positifs.

I - Grand O et petit o presque sûrement.

DEFINITIONS.

(a) La suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ est dite être un o (lire le nom de la lettre o) de a_n presque sûrement quand $n \rightarrow +\infty$, noté par

$$X_n = o(a_n), a.s. \text{ quand } n \rightarrow +\infty,$$

si et seulement si

$$(1.1) \quad \lim_{n \rightarrow +\infty} X_n/a_n = 0 \text{ a.s.}$$

(b) La suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ est dite être un grand O de a_n presque sûrement quand $n \rightarrow +\infty$, noté par

$$X_n = O(a_n), a.s. \text{ quand } n \rightarrow +\infty,$$

si et seulement si la suite $\{|X_n|/a_n, n \geq 1\}$ est presque sûrement bornée, c'est à dire

$$(1.2) \quad \lim_{n \rightarrow +\infty} \sup |X_n|/a_n < +\infty, a.s.$$

ATTENTION. Les signes d'égalité utilisés dans (1.1) et (1.2) doivent être lus dans une direction seulement, dans le sens suivant : Le membre de gauche est un petit o de a_n ou un grand O de a_n . N'inversons jamais l'égalité de la gauche à la droite. Par exemple, si X_n est un $o(n)$, c'est aussi un $o(n^2)$ et nous pouvons écrire $o(n) = o(n^2)$ *p.s.*, mais nous ne pouvons pas écrire $o(n^2) = o(n)$ *p.s.* Un exemple : $X_n = n^{3/2}$ est un $o(n^2)$ mais n'est pas un $o(n)$. Cette remarque s'étendra aux notations des petits o et des grands O en probabilité qui sont définie ci-dessous.

Cas particuliers concernant les constantes. Si $a_n = C > 0$ pour tout $n \geq 1$, noté $a_n \equiv C$, nous avons :

(i) $X_n = O(C)$ *p.s.* si et seulement si X_n/C est borné *p.s.* si et seulement si X_n est borné *p.s.* et nous écrivons

$$X_n = O(1) \text{ p.s.}$$

(ii) $X_n = o(C)$ *p.s.* si et seulement si $X_n/C \rightarrow 0$ *p.s.* si et seulement si $X_n/C \rightarrow 0$ *p.s.* et nous écrivons

$$X_n = o(1) \text{ p.s.}$$

(iii) Pour toute constante $C > 0$, nous écrivons $C = O(1)$.

PROPRIETES.

Les propriétés sont très nombreuses et les utilisateurs doivent souvent vérifier celle dépendant de ses travaux. Mais certaines d'entre elles doivent être connues et prêtes à être utilisées. Classons les en trois groupes.

Groupe A. Propriétés du petit o .

(1) $o(a_n)o(b_n) = o(a_nb_n)$ *p.s.*

(2) (1) $o(o(a_n)) = o(a_n)$ *p.s.*

(3) Si $b_n \geq a_n$ pour tout $n \geq 1$, $o(a_n) = o(b_n)$ *p.s.*

(4) $o(a_n) + o(a_n) = o(a_n)$ *p.s.*

(5) $o(a_n) + o(b_n) = o(a_n + b_n)$ *a.s.* et $o(a_n) + o(b_n) = o(a_n \vee b_n)$ *p.s.* où $a_n \vee b_n = \max(a_n, b_n)$.

(6) $o(a_n) = a_n o(1)$ *a.s.* et $a_n o(1) = O(a_n)$ *a.s.*

PREUVES. Chacune de ces propriétés est rapidement prouvée dans :

(1) Si $X_n = o(a_n)$ et $Y_n = o(b_n)$, alors

$$\lim_{n \rightarrow +\infty} \frac{|X_n Y_n|}{a_n b_n} = \lim_{n \rightarrow +\infty} \frac{|X_n|}{a_n} \times \lim_{n \rightarrow +\infty} \frac{|Y_n|}{b_n} = 0 \text{ p.s.}$$

et alors $X_n Y_n = o(a_n b_n)$ *p.s.*

(2) Si $Y_n = o(a_n)$, *p.s.* et $X_n = o(Y_n)$, *p.s.*,

$$\lim_{n \rightarrow +\infty} \frac{|X_n|}{a_n} = \lim_{n \rightarrow +\infty} \left| \frac{X_n}{Y_n} \right| \times \frac{|Y_n|}{a_n} = \lim_{n \rightarrow +\infty} \left| \frac{X_n}{Y_n} \right| \times \lim_{n \rightarrow +\infty} \frac{|Y_n|}{a_n} = 0,$$

c'est à dire $X_n = o(a_n)$ *p.s.*

(3) Si $X_n = o(a_n)$ et $b_n \geq a_n$ pour tout $n \geq 1$, alors

$$0 \leq \lim_{n \rightarrow +\infty} \sup \frac{|X_n|}{b_n} = \lim_{n \rightarrow +\infty} \sup \frac{|X_n| a_n}{a_n b_n} \leq \lim_{n \rightarrow +\infty} \sup \frac{|X_n|}{a_n} = 0 \text{ p.s.}$$

et $|X_n/b_n| \rightarrow 0$ *p.s.*, c'est à dire $X_n = o(b_n)$, *p.s.*

(4) Si $X_n = o(a_n)$ et $Y_n = o(a_n)$, alors

$$\lim_{n \rightarrow +\infty} \sup \frac{|X_n + Y_n|}{a_n} \leq \lim_{n \rightarrow +\infty} \frac{|X_n|}{a_n} + \lim_{n \rightarrow +\infty} \frac{|Y_n|}{b_n} = 0 \text{ p.s.}$$

et donc $X_n + Y_n = o(a_n)$ *p.s.*

(5) Pour prouver que $o(a_n) + o(b_n) = o(a_n + b_n)$ *p.s.*, utiliser le Point (3) pour voir que $o(a_n) = o(a_n + b_n)$ *p.s.* puisque $a_n + b_n \geq a_n$ pour tout $n \geq 1$ et aussi bien que $o(b_n) = o(a_n + b_n)$ *p.s.* et alors utiliser le Point (3) pour conclure. Nous prouvons que $o(a_n) + o(b_n) = o(a_n \vee b_n)$ *p.s.* de la même manière.

(6) C'est une simple utilisation de définition.

Groupe B. Propriétés du grand O.

(1) $O(a_n)O(b_n) = O(a_nb_n)$ *p.s.*

(2) $O(O(a_n)) = O(a_n)$ *p.s.*

(3) Si $b_n \geq a_n$ pour tout $n \geq 1$, $O(a_n) = O(b_n)$ *p.s.*

(4) $O(a_n) + O(a_n) = O(a_n)$ *p.s.*

(5) $O(a_n) + O(b_n) = O(a_n + b_n)$ *p.s.* et $O(a_n) + O(b_n) = O(a_n \vee b_n)$ *p.s.* où $a_n \vee b_n = \max(a_n, b_n)$.

(6) $O(a_n) = a_n O(1)$ *p.s.*, et $a_n O(1) = O(a_n)$ *p.s.*

PREUVES. Ces propriétés sont prouvées exactement comme celles du **Groupe A**, où les limites sont utilisées à la place des limites supérieures.

Groupe C. Propriétés des combinaisons des petits o's et grands O's.

(1) $o(a_n)O(b_n) = o(a_nb_n)$ *p.s.*

(2) $o(O(a_n)) = o(a_n)$ *p.s.* et $O(o(a_n)) = o(a_n)$, *p.s.*

(3a) Si $a_n = O(b_n)$, *p.s.*, alors $o(a_n) + O(b_n) = O(b_n)$ *p.s.*

(3b) Si $b_n = O(a_n)$, *p.s.*, alors $o(a_n) + O(b_n) = O(a_n)$ *p.s.*

(3c) Si $b_n = o(a_n)$, $p.s.$, alors $o(a_n) + O(b_n) = o(b_n)$ $p.s.$

(4) $(1 + o(a_n))^{-1} - 1 = o(a_n)$, $p.s.$

PREUVES.

(1) Si $X_n = o(a_n)$ et $Y_n = O(b_n)$, alors

$$\lim_{n \rightarrow +\infty} \sup \frac{|Y_n|}{b_n} = C < +\infty \text{ } p.s.$$

et

$$\begin{aligned} \lim_{n \rightarrow +\infty} \sup \frac{|X_n Y_n|}{a_n b_n} &= \lim_{n \rightarrow +\infty} \left(\left| \frac{X_n}{a_n} \right| \times \frac{|Y_n|}{b_n} \right) = \lim_{n \rightarrow +\infty} \sup \left| \frac{X_n}{a_n} \right| \times \lim_{n \rightarrow +\infty} \sup \frac{|Y_n|}{b_n} \\ &\leq C \lim_{n \rightarrow +\infty} \sup \left| \frac{X_n}{a_n} \right| = 0 \text{ } p.s. \end{aligned}$$

(2) Utiliser les Points (6) des Groupes A et B pour dire

$$o(O(a_n)) = o(1) \times O(a_n) = a_n \times o(1) \times O(1) = a_n \times o(1) = o(a_n)$$

et

$$O(o(a_n)) = o(a_n)O(1) = a_n \times o(1) \times O(1) = a_n \times o(1) = o(a_n).$$

(3a-b-c) Ces trois points sont prouvés de manières similaires. Donnons les détails de (3b) par exemple. Soient $X_n = o(a_n)$ et $Y_n = O(b_n)$ et $b_n = O(a_n)$. On a

$$\begin{aligned} o(a_n) + O(b_n) &= o(a_n) + O(O(a_n)) = o(a_n) + O(a_n) \\ &= a_n(o(1) + O(1)) = a_n \times O(1) = O(a_n). \end{aligned}$$

(4) Nous avons

$$\begin{aligned} (1 + o(a_n))^{-1} - 1 &= \frac{o(a_n)}{1 + o(a_n)} = \frac{o(a_n)}{1 + o(a_n)} = o(a_n)O(1) \\ &= a_n o(1)O(1) = a_n o(1) = a_n o(a_n). \end{aligned}$$

II - Grand O et petit o en probabilité.

DEFINITIONS.

(a) La suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ est dite être un o (lire le nom de la lettre o) de a_n en probabilité quand $n \rightarrow +\infty$, noté par

$$X_n = o_{\mathbb{P}}(a_n), \text{ quand } n \rightarrow +\infty,$$

si et seulement si

$$\lim_{n \rightarrow +\infty} X_n/a_n = 0$$

en probabilité,

c'est à dire pour tout $\lambda > 0$

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = 0.$$

(b) La suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ est dite être un grand O de a_n en probabilité quand $n \rightarrow +\infty$, noté par

$$X_n = O_{\mathbb{P}}(a_n), \text{ quand } n \rightarrow +\infty,$$

si et seulement si la suite $\{|X_n|/a_n, n \geq 1\}$ est bornée en probabilité, c'est à dire : Pour tout $\varepsilon > 0$, il existe une constante $\lambda > 0$, telle que

$$(1.3) \quad \inf_{n \geq 1} \mathbb{P}(|X_n| \leq \lambda a_n) \geq 1 - \varepsilon$$

ce qui est équivalente à

$$(1.4) \quad \liminf_{\lambda \uparrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = 0.$$

Avant d'aller plus loin, prouvons ceci :

LEMME 7. *Chacune des (1.3) et (1.4) est équivalente à : Pour tout $\varepsilon > 0$, il existe un entier $N \geq 1$ une constante $\lambda > 0$, telle que*

$$(1.5) \quad \inf_{n \geq N} \mathbb{P}(|X_n| \leq \lambda a_n) \geq 1 - \varepsilon.$$

PREUVES. Pour prouver cela pour (1.3) et (1.5), il suffira de prouver que (1.5) \implies (1.3) puisque l'implication inverse est évidente. Supposons que (1.5), c'est à dire pour $\varepsilon > 0$, ils existent $N \geq 1$ et un nombre réel $\lambda_0 > 0$ tels que

$$(1.6) \quad \forall (n \geq N), \mathbb{P}(|X_n/a_n| \leq \lambda_0) \geq 1 - \varepsilon.$$

Si $N = 1$, alors (1.3) est vérifiée. Si non, nous avons $j \in \{1, \dots, N-1\}$ fixé, $(|X_j/a_j| \leq \lambda) \uparrow \Omega$ quand $\lambda \uparrow +\infty$. Donc par le Théorème de la Convergence Monotone, il existe pour chaque $j \in \{1, \dots, N-1\}$ un nombre réel $\lambda_j > 0$ tel que $\mathbb{P}(|X_j/a_j| \leq \lambda_j) > 1 - \varepsilon$. Nous prenons $\lambda = \max(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ pour obtenir

$$\forall (n \geq 1), \mathbb{P}(|X_n/a_n| \leq \lambda) \geq 1 - \varepsilon,$$

ce qui est (1.3). Maintenant, prouvons que $(1.4) \iff (1.5)$. D'abord (1.4) veut dire

$$\lim_{\lambda \uparrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = 0,$$

puisque $\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n)$ est décroissante quand $\lambda \uparrow +\infty$ sur $[0, 1]$. Nous obtenons pour tout $\varepsilon > 0$, il existe un nombre réel $\lambda > 0$ tel que

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda a_n) = \lim_{N \uparrow +\infty} \sup_{n \geq N} \mathbb{P}(|X_n| > \lambda a_n) \leq \varepsilon/2.$$

Alors pour un quelconque $N > 0$,

$$\sup_{n \geq N} \mathbb{P}(|X_n| > \lambda a_n) \leq \varepsilon,$$

c'est à dire

$$\inf_{n \geq N} \mathbb{P}(|X_n| \leq \lambda a_n) \geq 1 - \varepsilon,$$

ce qui est (1.5). Qui à son tour, la reformulation de ceci donne : pour tout $\varepsilon > 0$, il existe $N_0 > 0$ et un nombre réel $\lambda_0 > 0$ tel que

$$(1.7) \quad \inf_{n \geq N} \mathbb{P}(|X_n| \leq \lambda_0 a_n) \geq 1 - \varepsilon,$$

c'est à dire

$$\sup_{n \geq N} \mathbb{P}(|X_n| > \lambda_0 a_n) < \varepsilon,$$

ce qui conduit à

$$\inf_{N \geq 1} \sup_{n \geq N} \mathbb{P}(|X_n| > \lambda_0 a_n) < \varepsilon,$$

et ensuite

$$\inf_{\lambda > 0} \inf_{N \geq 1} \sup_{n \geq N} \mathbb{P}(|X_n| > \lambda a_n) < \varepsilon,$$

ce qui est (1.4).

COMMENTAIRES, NOTATIONS ET QUELQUES LEMMES.

(a) A partir du Chapitre 3, un $O_{\mathbb{P}}(1)$ is simplement une suite de variables aléatoires tendues. A partir du Theorem 7 du Chapitre 3, nous avons que n'importe quelle suite $X_n = O_{\mathbb{P}}(a_n)$ contient une sous suite $(X_{n_k})_{k \geq 1}$ telle que $(X_{n_k}/a_{n_k})_{k \geq 1}$ converge vaguement dans \mathbb{R} .

(b) Il peut être convenant de reformuler (1.3) à la phrase suivante. Pour tout $\varepsilon > 0$, il existe un nombre réel $\lambda > 0$ tel que $|X_n| \leq \lambda a_n$

Avec une Probabilité Au Moins Egale à $1 - \varepsilon$ - noté $APAME(1 - \varepsilon)$ pour tout $n \geq 1$, ou $|X_n| > \lambda a_n$ Avec une Probabilité Au Plus Egale à ε - denoted $APAPE(\varepsilon)$ pour tout $n \geq 1$. (1.5), nous remplaçons (**pour** $n \geq 1$) par (**pour des grandes valeurs de** n) ou par (**pour n plus grand qu'un quelconque nombre** $n_0 > 1$).

Pour des démonstrations très longues, l'utilisation des types de phrases décrits ci dessus peut être utile.

Maintenant, nous pouvons donner quelques propriétés importantes des petits o 's et des grands O 's en probabilité.

Le lemme suivant est aussi utile.

LEMME 8. *Nous avons les propriétés suivantes*

(a) *Si X_n est une suite de k -vecteurs aléatoires convergeant en probabilité vers k -vecteur aléatoire X , alors $\|X_n\| = O_{\mathbb{P}}(1)$.*

(b) *Si X_n est une suite de vecteurs aléatoires à valeurs dans un espace métrique (S, d) convergeant en probabilité vers une constante $C \in S$ et si g est une application mesurable de (S, d) vers un espace métrique (E, r) . Ensuite si g est continue en C , alors $g(X_n)$ converge en probabilité vers C .*

(c) *Alors pour toute suite de k -vecteurs aléatoires $(X_n)_{n \geq 1}$ convergeant vers zero en probabilité. Soient $R(x)$ une fonction réelle de $x \in \mathbb{R}^k$ telle que $R(0) = 0$. Si $R(x) = o(\|x\|^p)$ quand $x \rightarrow 0$ pour un certain $p > 0$, alors $R(X_n) = o_{\mathbb{P}}(\|X_n\|^p)$. Si $R(x) = O(\|x\|^p)$ quand $x \rightarrow 0$ pour un certain $p > 0$, alors $R(X_n) = O_{\mathbb{P}}(\|X_n\|^p)$*

Preuve.

Preuve du Point (a). Si $X_n \rightarrow_{\mathbb{P}} X$, alors par la Proposition 12, $X_n \rightarrow_w X$ et par le Théorème 7 de l'application continue du Chapitre 2 $\|X_n\| \rightarrow_w \|X\|$. Alors par le Théorème 3, nous avons pour tout point de continuité de $F_{\|X\|}(\lambda) = P(\|X\| \leq \lambda)$,

$$\lim \mathbb{P}(|X_n| > \lambda) = \mathbb{P}(\|X\| > \lambda) = 1 - F_{\|X\|}(\lambda).$$

Puisque l'ensemble des points de continuité de $F_{\|X\|}$ est au plus dénombrable (see). Alors appliquer la formule ci dessus pour $\lambda \rightarrow +\infty$ lorsque les λ sont des points de continuité. Puisque $1 - F_{\|X\|}(\lambda) \rightarrow 0$

quand $\lambda \rightarrow +\infty$, alors pour tout $\varepsilon > 0$, nous sommes capable de choisir une valeur de $\lambda(\varepsilon)$ qui est un point de continuité de $F_{\|X\|}$ satisfaisant $1 - F_{\|X\|}(\lambda) < \varepsilon$. Pour tout $\varepsilon > 0$, nous avons trouvé $\lambda(\varepsilon) > 0$ tel que

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda(\varepsilon)) \leq \varepsilon,$$

ce qui implique

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda) = 0.$$

Le Point (a) est prouvé.

Preuve du Point (b). Supposons les notations de ce point et supposons que g est continue en C . Soit $\varepsilon > 0$. Par la continuité de g en C , il existe $\eta > 0$ tel que

$$d(x, C) < \eta \implies r(g(x), g(C)) < \varepsilon/2.$$

Maintenant

$$\begin{aligned} \mathbb{P}(r(g(X_n), g(C)) > \varepsilon) &= \mathbb{P}(\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) \geq \eta\}) \\ &+ \mathbb{P}(\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) < \eta\}) \\ &\leq \mathbb{P}(d(X_n, C) \geq \eta), \end{aligned}$$

puisque $(\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) < \eta\}) \subset (\{r(g(X_n), g(C)) > \varepsilon\} \cap \{d(X_n, C) < \eta\} \cap \{r(g(X_n), g(C)) < \varepsilon/2\}) = \emptyset$. Alors, puisque $X_n \rightarrow_{\mathbb{P}} C$, nous avons

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(r(g(X_n), g(C)) > \varepsilon) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(d(X_n, C) \geq \eta) = 0.$$

Ainsi le Point (b) est vrai.

Preuve du Point (c-1). Soit $R(x) = o(\|x\|^p)$ quand $x \rightarrow 0$. Alors $g(x) = |R(x)|/\|x\|^p \rightarrow 0$ quand $x \rightarrow 0$. Par la continuité de g en zéro et par le Point (a), $g(X_n) = |R(X_n)|/\|X_n\|^p \rightarrow_{\mathbb{P}} 0$ quand $n \rightarrow +\infty$ lorsque $X_n \rightarrow_{\mathbb{P}} X$ quand $n \rightarrow +\infty$. D'où $R(X_n) = o_{\mathbb{P}}(\|X_n\|^p)$.

Preuve du Point (c-2). Soit $R(x) = O(\|x\|^p)$ quand $x \rightarrow 0$. Alors pour tout $\varepsilon > 0$, ils existent $\eta > 0$ et $C > 0$ tel que $|R(x)|/\|x\|^p \leq C$ pour tout $\|x\| < \eta$. Alors pour $\lambda > C$,

$$\begin{aligned}
\mathbb{P}(|R(X_n)| / \|X_n\|^p > \lambda) &= \mathbb{P}(\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| \geq \eta\}) \\
&+ \mathbb{P}(\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| < \eta\}) \\
&\leq \mathbb{P}(\|X_n\| \geq \eta),
\end{aligned}$$

puisque $(\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| < \eta\}) \subset (\{|R(X_n)| / \|X_n\|^p > \lambda\} \cap \{\|X_n\| < \eta\} \cap \{|R(X_n)| / \|X_n\|^p < C\}) = \emptyset$. Alors pour tout $\lambda > C$,

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|R(X_n)| / \|X_n\|^p > \lambda) = 0$$

et ensuite

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|R(X_n)| / \|X_n\|^p > \lambda) = 0.$$

Nous concluons que $|R(X_n)| / \|X_n\|^p = O_{\mathbb{P}}(1)$.

PROPRIETES PRINCIPALES.

(1) Si $X_n = o(1)$ p.s., alors $X_n = o_{\mathbb{P}}(1)$.

(2) $o_{\mathbb{P}}(a_n) = a_n o_{\mathbb{P}}(1)$ et $a_n o_{\mathbb{P}}(1) = O_{\mathbb{P}}(a_n)$.

(3) $o_{\mathbb{P}}(a_n) o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n b_n)$.

(4) $o_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)$.

(5) Si $b_n \geq a_n$ pour tout $n \geq 1$, $o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(b_n)$.

(6) $o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(a_n)$.

(7) $o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n + b_n)$ et $o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n \vee b_n)$ p.s.
où $a_n \vee b_n = \max(a_n, b_n)$.

(8) Si $X_n = O(1)$ p.s., alors $X_n = O_{\mathbb{P}}(1)$.

(9) $O_{\mathbb{P}}(a_n) = a_n O_{\mathbb{P}}(1)$

(10) $O_{\mathbb{P}}(a_n) O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n b_n)$.

(11) $O_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = O_{\mathbb{P}}(a_n)$ a.s..

(12) Si $b_n \geq a_n$ pour tout $n \geq 1$, $O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(b_n)$.

(13) $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(a_n)$.

(14) $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n + b_n)$ et $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n \vee b_n)$ où $a_n \vee b_n = \max(a_n, b_n)$.

(15) $o_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_nb_n)$.

(16) $o_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)$ et $O_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)$.

(17a) Si $a_n = O_{\mathbb{P}}(b_n)$ alors $o(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(b_n)$ *p.s.*

(17b) Si $b_n = O_{\mathbb{P}}(a_n)$ alors $o_{<\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n)$.

(17c) Si $b_n = o_{\mathbb{P}}(a_n)$, *p.s.*, alors $o(a_n) + O_{\mathbb{P}}(b_n) = o(b_n)$.

(18) $(1 + o_{\mathbb{P}}(a_n))^{-1} - 1 = o_{\mathbb{P}}(a_n)$.

(19) Un $o_{\mathbb{P}}(1)$ est un $O_{\mathbb{P}}(1)$

PREUVES.

(1) Ceci vient de l'implication : $X_n \rightarrow 0$ *p.s.* $\implies X_n \rightarrow_P 0$ (Voir Proposition 12).

(2) Si $X_n = o_{\mathbb{P}}(a_n) \iff |X_n/a_n| \rightarrow_P 0 \iff X_n/a_n = o_{\mathbb{P}}(1) \iff X_n = a_n o_{\mathbb{P}}(1)$.

(3) Par le Point (2) ci dessus, $o_{\mathbb{P}}(a_n)o_{\mathbb{P}}(b_n) = a_nb_n \times o_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = a_nb_n \times o_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_nb_n)$ (Vérifions $o_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$ en Propriété (A1) dans la sous-section Appendice ci dessous).

(4) $o_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)o_{\mathbb{P}}(1) = a_n \times o_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = a_n \times o_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$ (Utilisons encore la Propriété (A1) de la sous-section Appendice ci dessous).

(5) Soit $b_n \geq a_n$ pour tout $n \geq 1$, $X_n = o_{\mathbb{P}}(a_n)$. Pour tout $\eta > 0$, $0 \leq \lim_{n \rightarrow +\infty} \sup P(|X_n/b_n| > \eta) \leq \lim_{n \rightarrow +\infty} P(|X_n/a_n| > \eta) = 0$.

(6) Soient $X_n = o_{\mathbb{P}}(a_n)$ et $Y_n = o_{\mathbb{P}}(a_n)$. Utiliser l'outil classique, pour $\eta > 0$,

$$\left(\frac{|X_n|}{a_n} > \eta/2 \right) \cap \left(\frac{|Y_n|}{a_n} > \eta/2 \right) \subset \left(\frac{|X_n + Y_n|}{a_n} > \eta \right).$$

Alors pour $\eta > 0$,

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|X_n + Y_n|}{a_n} > \eta \right) &\leq \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|X_n|}{a_n} > \eta/2 \right) \\ (1.8) \quad &+ \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\frac{|Y_n|}{a_n} > \eta/2 \right) = 0. \end{aligned}$$

(7) Pour prouver ce point, il suffit de combiner les Points (5) et (6) ci-dessus.

(8) $X_n = O(1)$ *a.s.* quand $n \rightarrow +\infty$ signifie qu'il existe Ω_0 mesurable telle que $\mathbb{P}(\Omega_0) = 1$ et pour tout $\omega \in \Omega_0$,

$$\limsup_{n \rightarrow +\infty} |X_n(\omega)| = \inf_{n \geq 1} \sup_{p \geq n} |X_p| = M(\omega) < +\infty.$$

Nous avons pour tout $n \geq 1$,

$$\mathbb{P}(|X_n| > \lambda) \leq \mathbb{P} \left(\sup_{p \geq n} |X_p| > \lambda \right)$$

Nous avons $Y_n = \sup_{p \geq n} |X_p| \downarrow M$ *p.s.* Alors $Y_n 1_{\Omega_0} \rightarrow_{\mathbb{P}} M 1_{\Omega_0}$. Nous travaillons avec les variables aléatoires et nous pouvons appliquer les résultats de la convergence vague. Nous obtenons $Y_n 1_{\Omega_0} \rightarrow_w M 1_{\Omega_0}$ par la Proposition 12. Par le Théorème 3, nous avons pour tout point de $F_M(\lambda) = P(M 1_{\Omega_0} \leq \lambda)$. Utilisons le Théorème de la Convergence Monotone pour obtenir

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda) &= \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{p \geq n} |X_p| 1_{\Omega_0} > \lambda \right) \\ &= \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{p \geq n} |X_p| 1_{\Omega_0} > \lambda \right) = \mathbb{P}(M 1_{\Omega_0} > \lambda) = 1 - F_M(\lambda). \end{aligned}$$

Puisque l'ensemble des points de discontinuité de F_M est au plus dénombrable (Voir). Nous appliquons alors la formule ci dessus pour $\lambda \rightarrow +\infty$ lorsque les λ sont des points de continuité. Puisque $1 - F_M(\lambda) \rightarrow 0$

quand $\lambda \rightarrow +\infty$, alors pour tout $\varepsilon > 0$, nous pouvons choisir une valeur de $\lambda(\varepsilon)$ qui est un point de continuité de F_M satisfaisant $1 - F_M(\lambda) < \varepsilon$. Pour tout $\varepsilon > 0$, nous trouver $\lambda(\varepsilon) > 0$ tel que

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda(\varepsilon)) \leq \varepsilon,$$

ce qui implique

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \lambda) = 0.$$

Donc $X_n = O_{\mathbb{P}}(1)$.

(9) Soit $X_n = O_{\mathbb{P}}(a_n)$. On a

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n/a_n| > \lambda) = 0.$$

C'est la définition que $X_n/a_n = O_{\mathbb{P}}(1)$ et alors $X_n = a_n O_{\mathbb{P}}(1)$.

(10) $O_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = a_n b_n O_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = a_n b_n O_{\mathbb{P}}(1)$ (Vérifions la Propriété ci dessous $O_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = O_{\mathbb{P}}(1)$ dans (A2) dans la sous-section Appendice).

(11) $O_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = O_{\mathbb{P}}(a_n)O_{\mathbb{P}}(1) = a_n O_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = O_{\mathbb{P}}(a_n)$.

(12) Soit $b_n \geq a_n$ pour tout $n \geq 1$ et $X_n = O_{\mathbb{P}}(a_n)$. On a

$$\lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n/b_n| > \lambda) \leq \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_n/a_n| > \lambda) = 0.$$

Donc $X_n = O_{\mathbb{P}}(b_n)$.

(13) Soit $X_n = O_{\mathbb{P}}(a_n)$ et $Y_n = O_{\mathbb{P}}(a_n)$. Utilisons la même technique comme dans la formule 1.8 ci-dessous pour trouver

$$\begin{aligned} \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\frac{|X_n + Y_n|}{a_n} > \lambda\right) &\leq \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\frac{|X_n|}{a_n} > \lambda/2\right) \\ (1.9) \quad &+ \lim_{\lambda \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}\left(\frac{|Y_n|}{a_n} > \lambda/2\right) = 0. \end{aligned}$$

(14) Combinons les Points (12) et (13) pour trouver celui-ci.

(15) $o_{\mathbb{P}}(a_n)O_{\mathbb{P}}(b_n) = a_nb_no_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = a_nb_no_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_nb_n)$. (Vérifions que $o_{\mathbb{P}}(1)O_{\mathbb{P}}(1)$ dans la propriété (A3) dans la sous-section Appendice ci dessous).

(16) $o_{\mathbb{P}}(O_{\mathbb{P}}(a_n)) = O_{\mathbb{P}}(a_n)o_{\mathbb{P}}(1) = a_nO_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = a_no_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$ et $O_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(a_n)$.

(17a) Soient $a_n = O(b_n)$ et $X_n = o_{\mathbb{P}}(a_n)$ et $Y_n = O_{\mathbb{P}}(b_n)$. Il existe $C > 0$ telle que $a_n \leq Cb_n$ pour tout $n \geq 1$. Alors $X_n = o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(Cb_n)$ par le Point (5). Mais évidemment $X_n = o_{\mathbb{P}}(Cb_n) = o_{\mathbb{P}}(b_n)$ et alors $X_n = O_{\mathbb{P}}(b_n)$ par le Point (19) ci dessous. Finalement $X_n + Y_n = O_{\mathbb{P}}(b_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(b_n)$.

(17b) Soient $b_n = O(a_n)$ et $X_n = o_{\mathbb{P}}(a_n)$ et $Y_n = O_{\mathbb{P}}(b_n)$. Nous échangeons les rôles de a_n et b_n pour obtenir $b_n \leq Ca_n$ et $Y_n = O_{\mathbb{P}}(Ca_n) = O_{\mathbb{P}}(a_n)$ par le Point (1) et finalement $X_n + Y_n = o_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(a_n) = O_{\mathbb{P}}(a_n)$.

(17c) Soient $b_n = o_{\mathbb{P}}(a_n)$ et $X_n = o_{\mathbb{P}}(a_n)$ et $Y_n = O_{\mathbb{P}}(b_n)$. Nous avons $X_n + Y_n = o_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(o_{\mathbb{P}}(a_n)) = o_{\mathbb{P}}(a_n) + o_{\mathbb{P}}(a_n)$ par le Point (16). Finalement $X_n + Y_n = o_{\mathbb{P}}(a_n)$.

(18) Nous avons

$$(1 + o_{\mathbb{P}}(a_n))^{-1} - 1 = \frac{o_{\mathbb{P}}(a_n)}{1 + o_{\mathbb{P}}(a_n)}.$$

Par le Point (b), $(1 + o_{\mathbb{P}}(a_n))^{-1} \rightarrow_P 1$ et par le Point (a) du même lemme, $(1 + o_{\mathbb{P}}(a_n))^{-1} = O_{\mathbb{P}}(1)$. Ainsi

$$(1 + o_{\mathbb{P}}(a_n))^{-1} - 1 = O_{\mathbb{P}}(1)o_{\mathbb{P}}(a_n) = o_{\mathbb{P}}(a_n),$$

par le Point (15).

(19) Par le Lemme 8, un $o_{\mathbb{P}}(1)$ converge vers O en probabilité et est alors un $O_{\mathbb{P}}(1)$.

1.1. Extensions. Les concepts de petit o et grand O dans \mathbb{R}^k de la façon suivante :

(1) La suite de vecteurs aléatoires $(X_n)_{n \geq 1}$ dans \mathbb{R}^k , est un $o(a_n)$ *p.s.* si et seulement si $\|X_n\|/a_n = o(1)$ *p.s.*, et est un $o_{\mathbb{P}}(a_n)$ si et seulement si $\|X_n\|/a_n = o_{\mathbb{P}}(1)$.

(b) La suite de vecteurs aléatoires $(X_n)_{n \geq 1}$ dans \mathbb{R}^k , est un $O(a_n)$ *p.s.* si et seulement si $\|X_n\|/a_n = O(1)$ *p.s.*, et est un $O_{\mathbb{P}}(a_n)$ si et seulement si $\|X_n\|/a_n = O_{\mathbb{P}}(1)$.

A partir de là, traiter ces concepts est facile en combinant leurs propriétés dans \mathbb{R} et celles des normes dans \mathbb{R}^k .

1.2. Suites équilibrées. Il peut aider dans certains cas d'avoir des suites X_n telles que $\|X_n\|$ et $1/\|X_n\|$ soient bornées en probabilité. Donnons quelques notations pour les suites réelles.

(1) Pour $0 \leq a < b < +\infty$, nous notons par $X_n = O_{\mathbb{P}}(a, b, a_n, b_n)$ la propriété que pour tout $\varepsilon > 0$, il existe $\lambda > 0$ tel que nous avons $(a + \lambda \leq |X_n|/a_n, |X_n|/a_n \leq b - \lambda)$ *APAME* $(1 - \varepsilon)$, pour des grandes valeurs de n . Si $a_n = b_n$ pour tout $n \geq 1$, nous écrivons simplement $X_n = O_{\mathbb{P}}(a, b, a_n)$.

(1) Pour $0 \leq a$, nous notons par $X_n = O_{\mathbb{P}}(a, +\infty, a_n, b_n)$ la propriété que pour tout $\varepsilon > 0$, il existe $\lambda > 0$ telle que nous avons $(a + \lambda^{-1} \leq |X_n|/a_n, |X_n|/a_n \leq \lambda)$ *APAME* $(1 - \varepsilon)$, pour des grandes valeurs de n . If $a_n = b_n$ pour tout $n \geq 1$, nous écrivons simplement $X_n = O_{\mathbb{P}}(a, +\infty, a_n)$.

Un exemple d'un $X_n = O_{\mathbb{P}}(0, +\infty, 1)$ est une suite X_n convergeant vaguement vers $X > 0$ *p.s.* Dans ce cas $1/X_n \rightarrow 1/X$ est finie *p.s.* et alors $X_n = O_{\mathbb{P}}(1)$ et $1/X_n = O_{\mathbb{P}}(1)$. Combinant ces deux points conduit à $X_n = O_{\mathbb{P}}(0, +\infty, 1)$.

1.3. Appendice. (A1) Si $X_n \rightarrow_{\mathbb{P}} a \in \mathbb{R}$ et $X_n \rightarrow_{\mathbb{P}} b \in \mathbb{R}$, alors $X_n Y_n \rightarrow_{\mathbb{P}} ab$.

Preuve. Nous avons $(\eta + |b|)\eta + |a|\eta \rightarrow 0$ quand $\eta \rightarrow 0$. pour tout $\varepsilon > 0$, pour tout $\delta > 0$, choisissons une valeur de $\eta > 0$ telle que $(\eta + |b|)\eta + |a|\eta < \delta$. Nous appliquons la définition des convergences $X_n \rightarrow_{\mathbb{P}} a$ et $X_n \rightarrow_{\mathbb{P}} b$ pour obtenir une valeur $N_0 \geq 1$ telle que pour tout $n \geq N_0$,

$$\mathbb{P}(|X_n - a| \geq \eta) \leq \varepsilon/2 \text{ et } \mathbb{P}(|Y_n - b| \geq \eta) \leq \varepsilon/2.$$

Mais

$$\begin{aligned} |X_n Y_n - ab| &= |X_n Y_n - a Y_n + a Y_n - ab| \\ &\leq |Y_n| |X_n - a| + |a| |Y_n - b| \\ &\leq (|Y_n - b| + |b|) |X_n - a| + |a| |Y_n - b| \end{aligned}$$

Sur $(|X_n - a| \geq \eta)^c \cap (|Y_n - b| \geq \eta)^c$,

$$|X_n Y_n - ab| \leq (\eta + |b|)\eta + |a|\eta \leq \delta.$$

D'où pour $n \geq N_0$,

$$(|X_n - a| \geq \eta)^c \cap (|Y_n - b| \geq \eta)^c \subset (|X_n Y_n - ab| \leq \delta),$$

c'est à dire

$$\mathbb{P}(|X_n - a| \geq \eta)^c \cap (|Y_n - b| \geq \eta)^c \leq \mathbb{P}(|X_n Y_n - ab| \leq \delta),$$

et en prenant les complémentaires,

$$\mathbb{P}(|X_n Y_n - ab| > \delta) \leq \mathbb{P}((|X_n - a| \geq \eta) \cup (|Y_n - b| \geq \eta)) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Par suite

$$X_n Y_n \rightarrow_{\mathbb{P}} ab.$$

Propriété (A2). $X_n = O_{\mathbb{P}}(1)$ et $X_n = O_{\mathbb{P}}(1)$ alors $X_n Y_n = O_{\mathbb{P}}(1)$.

Preuves. En appliquant la définition d'un $O_{\mathbb{P}}(1)$, nous pouvons trouver pour tout $\varepsilon > 0$, deux nombres entiers N_1 et N_2 et deux nombres positifs $\lambda_1 > 0$ et $\lambda_2 > 0$ tels que

$$\forall (n \geq N_1), \mathbb{P}(|X_n| \leq \lambda_1) \geq 1 - \varepsilon/2 \text{ et } \forall (n \geq N_2), \mathbb{P}(|Y_n| \leq \lambda_2) \geq 1 - \varepsilon/2.$$

Pour $n \geq \max(N_1, N_2)$,

$$(|X_n| \leq \lambda_1) \cap (|Y_n| \leq \lambda_2) \subset (|X_n Y_n| \leq \lambda_1 \lambda_2)$$

ce qui est équivalent à

$$(|X_n Y_n| > \lambda_1 \lambda_2) \subset (|X_n| > \lambda_1) \cup (|Y_n| > \lambda_2)$$

ce qui implique pour $n \geq \max(N_1, N_2)$,

$$\mathbb{P}(|X_n Y_n| > \lambda_1 \lambda_2) \leq \mathbb{P}(|X_n| > \lambda_1) + \mathbb{P}(|Y_n| > \lambda_2) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Ainsi pour tout $\varepsilon > 0$, il existe un nombre positif N ($= \max(N_1, N_2)$), il existe $\lambda > 0$ ($= \lambda_1 \lambda_2$) et pour tout $n \geq N$,

$$\mathbb{P}(|X_n Y_n| \leq \lambda) \geq 1 - \varepsilon..$$

D'où $X_n Y_n = O_{\mathbb{P}}(1)$.

$$X_n Y_n \rightarrow_{\mathbb{P}} ab.$$

Propriété (A3). Si $X_n = o_{\mathbb{P}}(1)$ et $Y_n = O_{\mathbb{P}}(1)$ alors $X_n Y_n = o_{\mathbb{P}}(1)$.

Preuve. Fixons $\varepsilon > 0$. En appliquant la définition d'un $O_{\mathbb{P}}(1)$ ils existent un nombre entier N_1 et un nombre positif $\lambda > 0$ tels que

$$\mathbb{P}(|Y_n| \leq \lambda) \geq 1 - \varepsilon/2.$$

Maintenant soit $\eta > 0$. Appliquons la définition de $X_n = o_{\mathbb{P}}(1)$ pour obtenir qu'il existe un entier positif N_2 tel que

$$\forall (n \geq N_2), \mathbb{P}(|X_n| > \eta/\lambda) \leq \varepsilon/2.$$

donc pour $n \geq \max(N_1, N_2)$,

$$(|X_n| \leq \eta/\lambda) \cap (|Y_n| \leq \lambda) \subset (|X_n Y_n| \leq \eta)$$

ce qui est équivalent à

$$(|X_n Y_n| > \eta) \subset (|X_n| > \eta/\lambda) \cup (|Y_n| > \lambda)$$

ce qui implique pour $n \geq \max(N_1, N_2)$,

$$\mathbb{P}(|X_n Y_n| > \eta) \leq \mathbb{P}(|X_n| > \eta/\lambda) + \mathbb{P}(|Y_n| > \lambda) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Ainsi pour tout $\varepsilon > 0$, pour tout $\eta > 0$, il existe un nombre positif N ($= \max(N_1, N_2)$), pour tout $n \geq N$,

$$\mathbb{P}(|X_n Y_n| > \eta) \leq \varepsilon.$$

D'où $X_n Y_n = o_{\mathbb{P}}(1)$.

2. Méthodes Delta

La méthode Delta, dans ses différentes versions, constitue un moyen rapide pour obtenir de nouvelles lois asymptotiques à partir d'une suite de variables aléatoires déjà convergente vaguement et définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Nous présenterons ici les versions univariées, vectorielles et matricielles. Nous verrons ici encore l'utilité des résultats de la section 6 du chapitre 2 lorsqu'ils sont combinés avec les règles de manipulations des petits o 's et des grands O 's en probabilité exposés dans la première section de ce chapitre.

Commençons avec la méthode Delta univariée, sur \mathbb{R} .

2.1. Version univariée de la méthode Delta.

PROPOSITION 33. *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et soit θ un nombre réel et $(a_n > 0)_{n \geq 1}$ une suite de nombres réels vérifiant $a_n \rightarrow +\infty$ as $n \rightarrow +\infty$.*

Soit $g : D \rightarrow \mathbb{R}$ une fonction de classe C^1 , telle que D soit un domaine de \mathbb{R} , que θ soit dans l'intérieur $\overset{\circ}{D}$ de D , et que l'on ait $\{X_n, n \geq 1\} \subset \overset{\circ}{D}$.

Si $a_n(X_n - \theta)$ converge vaguement vers une variable aléatoire Z quand $n \rightarrow +\infty$, alors

$$a_n(g(X_n) - g(\theta)) \rightsquigarrow g'(\theta)Z \quad \text{quand } n \rightarrow +\infty,$$

où $\nabla g(a) = g'(a)$ est la dérivée g en a .

Preuve de la proposition 33. Supposons que toutes les hypothèses de la proposition soient vraies. Par le point (a) du lemme 8, nous avons $a_n(X_n - \theta) = O_P(1)$ et alors,

$$X_n = \theta + O_P(1)a_n^{-1} \rightarrow_{\mathbb{P}} \theta.$$

Ceci, en vertu de la proposition 14 de la section 6 du chapitre 2, est équivalent à la convergence vague suivante

$$X_n \rightsquigarrow \theta.$$

Maintenant, par le théorème des accroissements finis, nous avons

$$(2.1) \quad g(X_n) - g(\theta) = g'(Y_n)(X_n - \theta),$$

où

$$\min(X_n, \theta) \leq Y_n \leq \max(X_n, \theta),$$

i.e.,

$$|Y_n - \theta| \leq |X_n - \theta|.$$

Il vient que $Y_n \rightarrow_{\mathbb{P}} \theta$ et puisque g' est continu, nous avons $g'(Y_n) \rightarrow_{\mathbb{P}} g'(\theta)$ en vertu du point (b) du lemme 8. Dès lors, en utilisant la proposition 14 de la section 6 du chapitre refcv, nous voyons que ceci est équivalent à

$$g'(Y_n) \rightsquigarrow g'(\theta).$$

Par la propriété de Slutsky donnée dans la formule 16 de la section section 6 du chapitre 2, nous avons la convergence vague suivante

$$(g'(Y_n), a_n(X_n - \theta)) \rightsquigarrow (g'(\theta), Z)$$

et par suite, en vertu du théorème de la transformation continue 7 du chapitre 2, combinée avec (2.1), nous obtenons le résultat final

$$a_n(g(X_n) - g(\theta)) = (g'(Y_n) \times a_n(X_n - \theta)) \rightsquigarrow g'(\theta)Z.$$

Remarque. Utilisons, non pas le nombre dérivé, mais plutôt la fonction (linéaire) dérivée définie par

$$h \rightarrow g'_\theta(h) = g'(\theta)h,$$

dans la proposition 33. Nous pouvons écrire alors le résultat sous la forme

$$a_n(g(X_n) - g(\theta)) = g'_\theta(Z).$$

Cette expression suggère la possibilité d'avoir ce résultats en dimension supérieurs and plus tard dans des espaces fonctionnels. Pour l'instant, passons aux formes multivariées.

2.2. Versions Multivariées. Le premier résultat concerne la transformation de la convergence vague d'une suite de suites variables aléatoires de k composantes par une fonction réelle de k arguments. Ceci donne la version vectorielle de la méthode Delta.

PROPOSITION 34. *Soit $(X_n)_{n \geq 1}$ une suite de vecteurs aléatoires de dimension k , $k \geq 1$, définis sur le même espace de probabilité (Ω, A, \mathbb{P}) et soit $\theta \in \mathbb{R}^k$ et $(a_n > 0)_{n \geq 1}$ une suite de nombres aléatoires tels que $a_n \rightarrow +\infty$ quand $n \rightarrow +\infty$.*

Soit $g : D \rightarrow \mathbb{R}$ une fonction de classe C^1 , telle que D soit un domaine de \mathbb{R}^k , que θ soit $\overset{\circ}{D}$ soit l'intérieur de D , et que l'on ait $\{X_n, n \geq 1\} \subset \overset{\circ}{D}$.

Si $a_n(X_n - \theta)$ converge vaguement vers un vecteur aléatoire Z de dimension k , lorsque $n \rightarrow +\infty$, alors nous avons

$$a_n(g(X_n) - g(\theta)) \rightsquigarrow {}^t\nabla g(\theta)Z = \langle \nabla g(\theta), Z \rangle \text{ as } n \rightarrow +\infty,$$

où

$${}^t\nabla g(\theta) = \left(\frac{\partial g(\theta)}{\partial \theta_1}, \dots, \frac{\partial g(\theta)}{\partial \theta_k} \right)$$

est le vecteur gradient de la fonction g en θ .

Le deuxième résultat concerne la transformation de la convergence vague d'une suite de suites variables aléatoires de k composantes par une fonction de m composantes de k arguments. Ceci donne la version matricielle de la méthode Delta.

PROPOSITION 35. *Soit $(X_n)_{n \geq 1}$ une suite de vecteurs aléatoires de dimension k , $k \geq 1$, définis sur le même espace de probabilité (Ω, A, \mathbb{P}) et soit $\theta \in \mathbb{R}^k$ et $(a_n > 0)_{n \geq 1}$ une suite de nombres aléatoires tels que $a_n \rightarrow +\infty$ quand $n \rightarrow +\infty$.*

Soit $g : D \rightarrow \mathbb{R}^m$ une fonction de classe C^1 , telle que D soit un domaine de \mathbb{R}^k , que θ soit $\overset{\circ}{D}$ soit l'intérieur de D , et que l'on ait $\{X_n, n \geq 1\} \subset \overset{\circ}{D}$. Notons par g_j , $1 \leq j \leq m$, les composantes de la fonction g .

Si $a_n(X_n - \theta)$ converge vaguement vers un vecteur aléatoire Z de dimension k , lorsque $n \rightarrow +\infty$, alors nous avons

$$a_n(g(X_n) - g(\theta)) \rightsquigarrow g'_\theta Z = \text{as } n \rightarrow +\infty,$$

où g'_θ est la matrice des dérivées partielles de premier ordre

$$g'_\theta = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \cdots & \frac{\partial g_1}{\partial \theta_j} & \cdots & \frac{\partial g_1}{\partial \theta_k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial g_i}{\partial \theta_1} & \cdots & \frac{\partial g_i}{\partial \theta_j} & \cdots & \frac{\partial g_i}{\partial \theta_k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial \theta_1} & \cdots & \frac{\partial g_m}{\partial \theta_j} & \cdots & \frac{\partial g_m}{\partial \theta_k} \end{pmatrix},$$

Preuve de la proposition 34. Supposons que les hypothèses de la proposition aient lieu.

Utilisons le développement de la fonction g au premier ordre en $\theta = {}^t(\theta_1, \dots, \theta_k)$ au point $x = {}^t(x_1, \dots, x_k)$:

$$(2.2) \quad g(x) - g(\theta) = (x_1 - \theta_1) \frac{\partial g}{\partial \theta_1}(\theta) + \dots + (x_k - \theta_k) \frac{\partial g}{\partial \theta_k}(\theta) + o(\|x - \theta\|).$$

Puisque $a_n(T_n - \theta) = a_n^t((T_{1,n}, \dots, T_{k,n}) - {}^t(\theta_1, \dots, \theta_k)) \rightsquigarrow Z = {}^t(Z_1, \dots, Z_k)$, il suit de l'application du théorème de la transformation continue 7 du chapitre 2, que pour tout $1 \leq i \leq k$, $a_n(T_{j,n} - \theta_j)$ converge vaguement vers Z_j et alors par le point (a) du lemme 8, nous avons

$$1 \leq j \leq k, (T_{j,n} - \theta_j) = O_{\mathbb{P}}(a_n^{-1}).$$

Par les points (10) and (13) des propriétés principales de la partie II de la section précédente, nous obtenons

$$(2.3) \quad \|T_n - \theta\| = \left\{ \sum_{j=1}^k (T_{j,n} - \theta_j)^2 \right\}^{1/2} = O_{\mathbb{P}}(a_n^{-1}) = o_{\mathbb{P}}(1).$$

Le terme $o(\|x - \theta\|)$ dans 2.2 est continu en tant que différence de deux fonctions continues et prend la valeur 0 pour $\|x - \theta\| = 0$. En utilisant la partie (c) du lemme 8, et en combinant cela avec (2.2) et (2.3), nous arrivons à

$$\begin{aligned} a_n(g(x) - g(\theta)) &= a_n(T_{1,n} - \theta_1) \frac{\partial g}{\partial \theta_1}(\theta) + \dots + a_n(T_{k,n} - \theta_k) \frac{\partial g}{\partial \theta_k}(\theta) + a_n o_{\mathbb{P}}(O_{\mathbb{P}}(a_n^{-1})) \\ &= {}^t \nabla g(\theta) (a_n(T_n - \theta) + o_{\mathbb{P}}(1)). \end{aligned}$$

Cela implique que $a_n(g(x) - g(\theta))$ et ${}^t\nabla g(\theta)(a_n(T_n - \theta))$ sont équivalents en probabilité. Puisque ${}^t\nabla g(\theta)(a_n(T_n - \theta)) \rightsquigarrow {}^t\nabla g(\theta)Z$ en vertu du théorème de la transformation 7 of Chapter 2, l'application de la proposition 15 de la Section 6 du chapitre 2, mène à

$$a_n(g(x) - g(\theta)) \rightsquigarrow {}^t\nabla g(\theta)Z.$$

Ceci finit la preuve de la proposition 35.

Preuve de la proposition 35. Assume that the hypotheses of the proposition hold.

La fonction g possède m composantes $g_j \in \mathbb{R}^m$ et nous pouvons écrire $g = {}^t(g_1, \dots, g_m)$. Chaque composante est de classe C^1 . Utilisons la conclusion de la proposition 34 pour chaque composante en $\theta = {}^t(\theta_1, \dots, \theta_k)$ for $x = {}^t(x_1, \dots, x_k)$ pour obtenir

$$(2.4) \quad g_j(x) - g_j(\theta) = (x_1 - \theta_1) \frac{\partial g_j}{\partial \theta_1}(\theta) + \dots + (x_k - \theta_k) \frac{\partial g_j}{\partial \theta_k}(\theta) + o(\|x - \theta\|).$$

Ecrivons cela sous forme matricielle ainsi qu'il suit

$$g(x) - g(\theta) = g'_\theta(x - \theta) + o^{(m)}(\|x - \theta\|),$$

ù $o^{(m)}(\|x - \theta\|)$ est un vecteur de m coordonnées tel que chaque composante est une fonction continue, et aussi un $o(\|x - \theta\|)$. Une notation similaire est aussi utilisée pour $o_{\mathbb{P}}(\circ)$. En appliquant la méthode déjà utilisée dans la proposition 34, nous obtenons

$$g(T_n) - g(\theta) = g'_\theta(T_n - \theta) + o_{\mathbb{P}}^{(m)}(a_n^{-1})$$

et

$$a_n(g(T_n) - g(\theta)) = g'_\theta a_n(T_n - \theta) + o_{\mathbb{P}}^{(m)}(1).$$

Nous avons donc

$$\|a_n(g(T_n) - g(\theta)) - g'_\theta a_n(T_n - \theta)\|_{\mathbb{R}^m} = \|o_{\mathbb{P}}^{(m)}(1)\|_{\mathbb{R}^m} = o_{\mathbb{P}}(1).$$

Dès lors $a_n(g(T_n) - g(\theta))$ has the same weak limit as $g'_\theta a_n(T_n - \theta)$ which is $g'_\theta Z$ by the continuous mapping.

3. Utilisation du Processus Empirique Fonctionnel en Statistique Asymptotique

3.1. Processus Empirique Fonctionnel. Le Processus Empirique Fonctionnel (*PEF*) est un outil puissant qui peut être utilisé pour obtenir des distributions limites. Il est similaire à la méthode Delta. Cependant, la méthode du *PEF* possède un avantage que nous décrivons ci-bas.

Etant donnée une suite de variables aléatoires Z_1, Z_2, \dots , définies sur le même espace de probabilité, indépendantes et identiquement distribuées, de loi de probabilité commune \mathbb{P}_0 à valeurs dans un espace métrique S , que nous prenons égal \mathbb{R}^k , $k \geq 1$ ici. Nous serons en mesure

(1) de trouver un processus gaussien $\mathbb{G}_{\mathbb{P}_0}$ indexé par des fonctions f de S dans \mathbb{R} ,

et

(2) d'exprimer les distribution asymptotiques de statistiques qui sont des fonctions of Z_1, Z_2, \dots, Z_n par rapport au processus gaussien \mathbb{G}_0 .

Cette méthode a l'avantage énorme de pouvoir travailler sur toutes les statistiques fonctions de Z_1, Z_2, \dots, Z_n de manière modulaire. En effet, on peut travailler de manière séparée pour chacune de ces statistiques comme and un catalogue. Et à tout moment, on peut prendre un certain nombre de ces statistiques et obtenir immédiatement leur loi conjointe limite, grâce au champ gaussien $\mathbb{G}_{\mathbb{P}_0}$. **Nous dirons que ces dsitributions limites sont exprimées dans le champ gaussien de $\mathbb{G}_{\mathbb{P}_0}$.** Au contraire la méthode Delta ne possède pas cette souplesse modulaire. Dans chaque cas, il faut faire des calculs spécifiques.

Un autre fait à signaler est que les lois limites conjointes obtenues par la méthode du *PEF* utilisent des variances et covariances exprimées sous des formes fonctionnelles. Cela rend les calculs directs et fluides. A l'arrivée, ces variances peuvent sembler compliquées et non calculables à la main. Mais nous ne nous en occupons pas. Les ordinateurs puissants de notre ère sont l pour les calculer en des temps très courts. Cela implique que l'utilisation de cette méthode, dans beaucoup de cas,

doit s'accompagner avec l'écriture de codes dans des logiciels courants.

Avant de présenter le Processus Empirique Fonctionnel, nous voudrions rassurer le lecteur que nous n'utiliserons ici que les limites en distributions finies de ce processus, restant ainsi dans le cadre vectoriel de ce monograph. Nous ne ferons pas appel aux outils puissants de convergence uniforme ni aux classes de Vapnick-Cervonenkis.

Soit Z_1, Z_2, \dots une suite de copies indépendantes d'une variable aléatoire Z définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace métrique (S, d) . Définissons pour chaque $n \geq 1$, le Processus Empirique Fonctionnel (PEF) par

$$\mathbb{G}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(Z_i) - \mathbb{E}f(Z_i)),$$

où f est une fonction mesurable définie de \mathbb{R} dans

$$(3.1) \quad \mathbb{V}_Z(f) = \int (f(x) - \mathbb{P}_Z(f))^2 dP_Z(x) < \infty,$$

ce qui implique que

$$(3.2) \quad \mathbb{P}_Z(|f|) = \int |f(x)| dP_Z(x) < \infty.$$

Notons par $\mathcal{F}(S)$ - \mathcal{F} en court - la class des fonctions réelles définies sur S vérifiant (3.1). Cette espace \mathcal{F} , muni de l'addition et de la multiplication externe par des scalaires de \mathbb{R} , est un espace linéaire.

L'une des propriétés les plus importantes de \mathbb{G}_n est qu'elle est un opérateur linéaire sur \mathcal{F} , i.e., pour tout f et g éléments de \mathcal{F} et pour tout $(a, b) \in \mathbb{R}^2$, nous avons

$$a\mathbb{G}_n(f) + b\mathbb{G}_n(g) = \mathbb{G}_n(af + bg).$$

La convergence vague en distribution est la suivante

LEMME 9. *Given the notation above, then for any finite number of elements f_1, \dots, f_p of \mathcal{F} , $k \geq 1$, we have*

$${}^t(\mathbb{G}_n(f_1), \dots, \mathbb{G}_n(f_k)) \rightsquigarrow \mathcal{N}_k(0)$$

where

$$\Gamma(f_i, f_j) = \int (f_i - \mathbb{P}_Z(f_i)) (f_j - \mathbb{P}_Z(f_j)) d\mathbb{P}_Z(x), 1 \leq j \leq k.$$

Ce lemme nous dit que la limite vague de la suite de vecteurs ${}^t(\mathbb{G}_n(f_1), \mathbb{G}_n(f_2), \dots, \mathbb{G}_n(f_k))$ a la même loi que le vecteur ${}^t(\mathbb{G}(f_1), \mathbb{G}(f_2), \dots, \mathbb{G}(f_k))$, où $\{\mathbb{G}(f), f \in \mathcal{F}\}$ est un process gaussien de fonction variance-covariance

$$(3.3) \quad \Gamma(f, g) = \int (f - \mathbb{P}_Z(f)) (g - \mathbb{P}_Z(g)) d\mathbb{P}_Z(x), \quad (f, g) \in \mathcal{F}^2.$$

En appliquant le théorème de Skorohod-Wichura 5 (Voir Chapitre 2), nous pouvons supposer que nous sommes sur un espace de probabilité et que nous avons l'approximation suivante

$$(3.4) \quad \mathbb{G}_n(f_1) = \mathbb{G}_n(f_1) + o_{\mathbb{P}}(1), 1 \leq i \leq p.$$

Nous reviendrons sur l'application de cette formule.

Preuve du lemme 9. Il suffira d'appliquer le critère de Cramér-Wold (voir Proposition 1 in 1), en montrant que pour tout $a = {}^t(a_1, \dots, a_k) \in \mathbb{R}^k$, nous avons

$$\langle a, T_n \rangle \rightsquigarrow \langle a, T \rangle$$

où nous avons noté $T_n = {}^t(\mathbb{G}_n(f_1), \dots, \mathbb{G}_n(f_k))$ et où T suit une loi $\mathcal{N}_k(0, \Gamma(f_i, f_j)_{1 \leq i, j \leq k})$ et $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel de \mathbb{R}^k .

Mais, par le théorème central limite dans \mathbb{R} , nous avons

$$\langle a, T_n \rangle = \mathbb{G}_n \left(\sum_{i=1}^k a_i f_i \right) \rightsquigarrow \mathcal{N}(0, \sigma_{\infty}^2),$$

où, avec la notation $g = \sum_{1 \leq i \leq k} a_i f_i$, nous avons

$$\sigma_{\infty}^2 = \int (g(x) - \mathbb{P}_Z(g))^2 d\mathbb{P}_Z(x)$$

et ceci donne aisément

$$\sigma_{\infty}^2 = \sum_{1 \leq i, j \leq k} a_i a_j \Gamma(f_i, f_j),$$

si bien que $N(0, \sigma_\infty^2)$ est la loi de $\langle a, T \rangle$. La preuve est finie.

3.2. Comment utiliser l'outil du *PEF*? En général, les statistiques avec lesquelles nous travaillons en général utilisent des données univariées ou multivariées ou tics, c'est-à-dire que nous travaillons souvent dans \mathbb{R}^k . Une fois que nous avons notre échantillon Z_1, Z_2, \dots , constitué de vecteurs définis sur le même espace de probabilité et à valeurs dans \mathbb{R}^k , l'essentiel des statistiques étudiées sont de la forme

$$H_n = \frac{1}{n} \sum_{i=1}^k H(Z_i)$$

où $H \in \mathcal{F}$. Dans un tel cas, nous utilisons Lemma 9 et le point point (a) du lemme 8, pour avoir ce développement très simple $\mu(H) = \mathbb{E}H(Z)$,

$$(3.5) \quad H_n = \mu(H) + n^{-1/2} \mathbb{G}_n(H).$$

Nous avons que $\mathbb{G}_n(H)$ est asymptotiquement borné en probabilité puisque $\mathbb{G}_n(H)$ converge vaguement, disons vers $M(H)$, et par la suite, par le théorème de la transformation continue (see Theorem 7, Chapitre 2), nous aurons $\|\mathbb{G}_n(H)\| \rightsquigarrow \|M(H)\|$. Puisque toutes variables $\mathbb{G}_n(H)$ sont définies sur le même espace de probabilité, nous aurons en vertu de l'assertion du théorème Portmanteau (Theorem 2, Chapitre 2) relative aux ensembles ouverts, que pour tout $\lambda > 0$,

$$\limsup_{n \rightarrow \infty} P(\|\mathbb{G}_n(H)\| > \lambda) \leq P(\|M(H)\| > \lambda).$$

Par la suite, nous avons

$$\liminf_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\|\mathbb{G}_n(H)\| > \lambda) \leq \liminf_{\lambda \rightarrow \infty} \mathbb{P}(\|M(H)\| > \lambda) = 0.$$

A partir d'ici, nous faisons appel aux notation $O_{\mathbb{P}}$, en disant que nous avons : $\mathbb{G}_n(H) = O_{\mathbb{P}}(1)$. La formule (3.5) devient

$$H_n = \mu(H) + n^{-1/2} \mathbb{G}_n(H) = \mu(H) + O_{\mathbb{P}}(n^{-1/2})$$

A présent, utilisons la méthode delta. En effet, soit $g : \mathbb{R} \mapsto \mathbb{R}$ une fonction continûment différentiable dans un voisinage de $\mu(H)$. The théorème des accroissements finis mène

$$(3.6) \quad g(H_n) = g(\mu(H)) + g'(\mu_n(H)) n^{-1/2} \mathbb{G}_n(H)$$

où

$$\mu_n(H) \in [(\mu(H) + n^{-1/2}\mathbb{G}_n(H)) \wedge \mu(H), (\mu(H) + n^{-1/2}\mathbb{G}_n(H)) \vee \mu(H)]$$

si bien que

$$|\mu_n(H) - \mu(H)| \leq n^{-1/2}\mathbb{G}_n(H) = O_{\mathbb{P}}(n^{-1/2}).$$

Dès lors $\mu_n(H)$ converge vers $\mu(H)$ en probabilité (noté $\mu_n(H) \rightarrow_{\mathbb{P}} \mu(H)$). Mais la convergence en probabilité vers une constante équivaut à la convergence vague. Donc $\mu_n(H) \rightsquigarrow \mu(H)$. En utilisant encore le théorème de la transformation continue (see Theorem 7, Chapitre 2), nous obtenons $g'(\mu_n(H)) \rightsquigarrow g'(\mu(H))$. En ré-utilisant l'équivalence entre convergence vague et convergence en probabilité vers une constante, il vient que $g'(\mu_n(H)) \rightarrow_{\mathbb{P}} g'(\mu(H))$. Maintenant, (3.6) devient

$$\begin{aligned} g(H_n) &= g(\mu(H)) + (g'(\mu(H) + o_P(1)) n^{-1/2}\mathbb{G}_n(H) \\ &= g(\mu(H)) + g'(\mu(H) \times n^{-1/2}\mathbb{G}_n(H) + o_P(1)) n^{-1/2}\mathbb{G}_n(H) \\ &= g(\mu(H)) + n^{-1/2}\mathbb{G}_n(g'(\mu(H)H) + o_P(n^{-1/2})) \end{aligned}$$

Nous arrivons au développement final

$$(3.7) \quad g(H_n) = g(\mu(H)) + n^{-1/2}\mathbb{G}_n(g'(\mu(H)H) + o_P(n^{-1/2})).$$

En utilisant la représentation de Skorohod-Wichura, nous obtenons à travers la formule, que

$$(3.8) \quad g(H_n) = g(\mu(H)) + n^{-1/2}\mathbb{G}(g'(\mu(H)H) + o_P(n^{-1/2})).$$

La méthode consiste à utiliser le développement (3.7) autant de fois que possible, et de faire certains calculs algébriques sur eux.

Ces calculs algébriques mentionnés ci-haut consistent par ailleurs à appliquer les résultats du lemma ci-dessous.

LEMME 10. *Let (A_n) and (B_n) be two sequences of real valued random variables defined on the same probability space holding the sequence Z_1, Z_2, \dots*

Let A and B be two real numbers and Let $L(z)$ and $H(z)$ be two real-valued functions of $z \in S$, with $(L, H) \in \mathcal{F}^2$.

Suppose that

$$A_n = A + n^{-1/2}\mathbb{G}_n(L) + o_P(n^{-1/2})$$

and

$$A_n = B + n^{-1/2}\mathbb{G}_n(H) + o_P(n^{-1/2}).$$

Then

$$A_n + B_n = A + B + n^{-1/2}\mathbb{G}_n(L + H) + o_P(n^{-1/2}),$$

and

$$A_n B_n = AB + n^{-1/2}\mathbb{G}_n(BL + AH)$$

and if $B \neq 0$,

$$\frac{A_n}{B_n} = \frac{A}{B} + n^{-1/2}\mathbb{G}_n\left(\frac{1}{B}L - \frac{A}{B^2}H\right) + o_P(n^{-1/2})$$

Preuve. Cette preuve est donnée au lecteur en guise d'exercices.

En mettant ensemble toutes les étapes précédente de manière intelligente, la méthodologie aboutit à un résultat de la forme

$$\begin{aligned} T_n &= t + n^{-1/2}\mathbb{G}_n(h) + o_P(n^{-1/2}) \\ &= t + n^{-1/2}\mathbb{G}(h) + o_P(n^{-1/2}) \end{aligned}$$

qui entraîne la convergence vague

$$\begin{aligned} \sqrt{n}(T_n - t) &= \mathbb{G}_n(h) + o_P(1) \rightsquigarrow \mathbb{N}(0, \Gamma(h, h)) \\ &= \mathbb{G}(h) + o_P(1). \end{aligned}$$

Dans cette dernière partie, nous allons montrer comment appliquer cet outil à un cas non relatif au coefficient de corrélation linéaire.

3.3. Un exemple. Appliquons la méthode à l'estimateur plug-in coefficient de corrélation linéaire. Un estimateur plug-in d'une statistique est sa forme empirique. L'estimateur plug-in du coefficient de corrélation linéaire de deux variables aléatoires réelles (X, Y) , telle que ni X et ni Y ne sont dégénérées, est défini par

$$\rho = \frac{\sigma_{xy}}{\sigma_x^2 \sigma_y^2}$$

où

$$\mu_x = \int x dP_X(x), \mu_y = \int y dP_Y(y), \sigma_{xy} = \int (x - \mu_x)(y - \mu_y) dP_{(X,Y)}(x, y)$$

et

$$\sigma_x^2 = \int (x - \mu_x)^2 dP_X(x), \quad \sigma_y^2 = \int (y - \mu_y)^2 dP_Y(y).$$

Nous excluons aussi le cas où nous avons $|\rho| = 1$, cas dans lequel une des variables X et Y est une fonction affine de l'autre, c'est-à-dire par exemple que nous avons $X = aY + b$ pour $(a, b) \in \mathbb{R}^2$.

Il est clair que le centrage des variables X et Y en leur moyennes et leur normalisation par les écart-types σ_x et σ_y n'affecte pas le coefficient de corrélation linéaire ρ . Ainsi nous pouvons centrer et normaliser X et Y et par la suite, assumer que nous avons

$$\mu_x = \mu_y = 0, \quad \sigma_x = \sigma_y = 1.$$

Cependant, nous laisserons figurer les moyennes et écart-types. C'est seulement la conclusion que nous prendrons les valeurs particulières.

Construisons l'estimateur plug-in estimator de ρ . Pour cela, soit $(X_1, Y_1), (X_2, Y_2), \dots$ une suite d'observations indépendantes de observations of (X, Y) . Pour tout $n \geq 1$, l'estimateur plug-in est

$$\rho_n = \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right\} \left\{ \frac{1}{n^2} \sum_{i=1}^n (X_i - \bar{X})^2 \times \sum_{i=1}^n (Y_i - \bar{Y})^2 \right\}^{-1/2}.$$

Nous allons donner les propriétés asymptotiques de ρ_n en tant qu'estimateur de ρ . Introduisons les notations :

$$\mu_{(p,x),(q,y)} = E((X - \mu_x)^p (Y - \mu_y)^q), \quad \mu_{4,x} = E(X - \mu_x)^4, \quad \mu_{4,y} = E(Y - \mu_y)^4$$

L'application de la méthode donne le résultat suivant.

THÉORÈME 12. *Supposons ni X ni Y ne soit dégénérée et ont toutes leurs deux des moments finis jusqu' l'ordre 4, que X^3Y et XY^3 soit de moyenne finies. Alors, quand $n \rightarrow \infty$,*

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma^2),$$

où

$$\begin{aligned} \sigma^2 = & \sigma_x^{-2} \sigma_y^{-2} (1 + \rho^2/2) \mu_{(2,x),(2,y)} + \rho^2 (\sigma_x^{-4} \mu_{4,x} + \sigma_y^{-4} \mu_{4,y})/4 \\ & - \rho (\sigma_x^{-3} \sigma_y^{-1} \mu_{(3,x),(1,y)} + \sigma_x^{-1} \sigma_y^{-3} \mu_{(1,x),(3,y)}) \end{aligned}$$

Ce résultat permet de tester l'indépendance entre X et Y , ou de tester l'absence de corrélation linéaire dans le sens suivant.

THÉORÈME 13. *Supposons que les hypothèses du théorème 12 soient vraies. Alors, nous avons les assertions suivantes :*

(1) *Si X et Y ne sont pas linéairement corrélées, i.e. $\rho = 0$, nous avons*

$$\sqrt{n} \rho_n \rightsquigarrow \mathcal{N}(0, \sigma_1^2),$$

où

$$\sigma_1^2 = \sigma_x^{-2} \sigma_y^{-2} \mu_{(2,x),(2,y)}.$$

(2) *Si X et Y sont indépendantes entre elles, alors $\rho = 0$, et*

$$\sqrt{n} \rho_n \rightsquigarrow \mathcal{N}(0, 1)$$

Preuves. Nous allons utiliser le *PEF* basé sur les observations (X_i, Y_i) , $i = 1, 2, \dots$, qui sont des copies indépendantes de (X, Y) . Ecrivons

$$\rho_n^2 = \frac{\frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \bar{Y}}{\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2}} = \frac{A_n}{B_n}.$$

Nous disons une fois pour toute que toutes les fonctions de $Z = (X, Y)$ qui apparaissent dans les calculs sont mesurables et ont des seconds moments finis. Traitons séparément le numérateur et le dénominateur. l'utilisation du *PEF* pour A_n gives

$$(3.9) \quad \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i Y_i = \mu_{xy} + n^{-1/2} G_n(p), \\ \bar{X} = \mu_x + n^{-1/2} G_n(\pi_1), \\ \bar{Y} = \mu_y + n^{-1/2} G_n(\pi_2), \end{cases}$$

où $p(x, y) = xy$, $\pi_1(x, y) = x$, $\pi_2(x, y) = y$. A partir de là, nous utilisons le fait que $G_n(g) = O_P(1)$ pour $\mathbb{E}(g(X, Y)^2) < +\infty$ et obtenons

$$(3.10) \quad A_n = \mu_{xy} + n^{-1/2}G_n(p) - (\mu_x + n^{-1/2}G_n(\pi_1))(\mu_y + n^{-1/2}G_n(\pi_2)).$$

Cela mène à

$$A_n = \sigma_{xy} + n^{-1/2}G_n(H_1) + o_P(n^{-1/2})$$

avec

$$H_1(x, y) = p(x, y) - \mu_x \pi_2 - \mu_y \pi_1.$$

Maintenant, nous devons traiter B_n . Par symétrie des rôles de $\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2}$ et de $\left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2}$, nous traitons l'une de ces expressions et étendons les résultats l'autre. Travaillons avec $\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2}$. La combinaison de (3.9) et de la méthode Delta conduit à

$$\bar{X}^2 = (\mu_x + n^{-1/2}G_n(\pi_1))^2 = \mu_x^2 + 2\mu_x n^{-1/2}G_n(\pi_1) + o_P(n^{-1/2}),$$

c'est-à-dire,

$$\bar{X}^2 = (\mu_x + n^{-1/2}G_n(\pi_1))^2 = \mu_x^2 + n^{-1/2}G_n(2\mu_x \pi_1) + o_P(n^{-1/2}).$$

Nous obtenons

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 &= m_{2,x} + n^{-1/2}G_n(\pi_1^2) - \bar{X}^2 \\ &= m_{2,x} - \mu_x^2 + n^{-1/2}G_n(\pi_1^2 - 2\mu_x \pi_1) + o_P(n^{-1/2}) \\ &= \sigma_x^2 + n^{-1/2}G_n(\pi_1^2 - 2\mu_x \pi_1) + o_P(n^{-1/2}). \end{aligned}$$

Une nouvelle application de la méthode Delta donne

$$\left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2} = \sigma_x + n^{-1/2}G_n\left(\frac{1}{2\sigma_x} \{ \pi_1^2 - 2\mu_x \pi_1 \} \right) + o_P(n^{-1/2}).$$

Étendons les résultats sur les Y_i pour avoir

$$\left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2} = \sigma_y + n^{-1/2}G_n\left(\frac{1}{2\sigma_y} \{ \pi_2^2 - 2\mu_y \pi_2 \} \right) + o_P(n^{-1/2}).$$

Nous arrivons à

$$\begin{aligned} B_n &= \left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right\}^{1/2} \\ &= \sigma_x \sigma_y + n^{-1/2} G_n \left(\frac{\sigma_y}{2\sigma_x} \{ \pi_1^2 - 2\mu_x \pi_1 \} + \frac{\sigma_x}{2\sigma_y} \{ \pi_2^2 - 2\mu_y \pi_2 \} \right) + o_P(n^{-1/2}). \end{aligned}$$

En mettant

$$H_2(x, y) = \frac{\sigma_y}{2\sigma_x} \{ \pi_1^2 - 2\mu_x \pi_1 \} + \frac{\sigma_x}{2\sigma_y} \{ \pi_2^2 - 2\mu_y \pi_2 \},$$

nous avons

$$(3.11) \quad B_n = \sigma_x \sigma_y + n^{-1/2} G_n(H_2) + n^{-1/2}.$$

Maintenant, la combinaison de (3.10) et de (3.11), et l'utilisation du lemme 10 donne

$$\sqrt{n}(\rho_n^2 - \rho^2) = n^{-1/2} G_n \left(\frac{1}{\sigma_x \sigma_y} H_1 - \frac{\sigma_{xy}}{\sigma_x^2 \sigma_y^2} H_2 \right) + o_P(1).$$

Posons

$$H = \frac{1}{\sigma_x \sigma_y} (p(x, y) - \mu_x \pi_2 - \mu_y \pi_1) - \frac{\rho}{\sigma_x \sigma_y} \left\{ \frac{1}{2\sigma_x^2} \{ \pi_1^2 - 2\mu_x \pi_1 \} + \frac{1}{2\sigma_y^2} \{ \pi_2^2 - 2\mu_y \pi_2 \} \right\}.$$

Nous continuons avec les variables centrées et normalisées. Nous avons

$$H(x, y) = p(x, y) - \frac{\rho}{2} (\pi_1^2 + \pi_2^2)$$

et

$$H(X, Y) = XY - \frac{\rho}{2} (X^2 + Y^2).$$

Adoptons la notation

$$\mu_{(p,x),(q,y)} = E((X - \mu_x)^p (Y - \mu_y)^q).$$

Nous obtenons

$$\mathbb{E}H(X, Y) = \sigma_{xy} - \rho = 0$$

avec la remarque que $\text{var} H(X, Y)$ is égale à

$$\mu_{(2,x),(2,y)} + \rho^2(\mu_{4,x} + \mu_{4,y})/4 - \rho(\mu_{(3,x),(1,y)} + \mu_{(1,x),(3,y)}) + \rho^2 \mu_{(2,x),(2,y)}/2$$

et finalement

$$\text{var}H(X, Y) = \sigma_0^2$$

avec

$$\sigma_0^2 = (1 + \rho^2/2)\mu_{(2,x),(2,y)} + \rho^2(\mu_{4,x} + \mu_{4,y})/4 - \rho(\mu_{(3,x),(1,y)} + \mu_{(1,x),(3,y)}).$$

Ceci donne la conclusion pour des variables X et Y normalisées :

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma_0^2).$$

Enfin, lorsque nous utilisons les coefficients de normalisation dans in σ_0 , nous obtenons

$$\begin{aligned} \sigma^2 &= \sigma_x^2 \sigma_y^2 (1 + \rho^2/2) \mu_{(2,x),(2,y)} + \rho^2 (\sigma_x^4 \mu_{4,x} + \sigma_y^4 \mu_{4,y})/4 \\ &\quad - \rho (\sigma_x^3 \sigma_y \mu_{(3,x),(1,y)} + \sigma_x \sigma_y^3 \mu_{(1,x),(3,y)}) \end{aligned}$$

pour conclure dans le cas général que

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma^2)$$

La preuve du théorème 13 découle la suite calculs directs sous les conditions particulières de ρ et en présence de l'indépendance.

Théorie des fonctions et éléments d'analyse réelle à travers des exercices

1. Revue des limites dans $\overline{\mathbb{R}}$. Ce que nous ne devons pas ignorer des limites.

Definition: Un nombre réel $\ell \in \overline{\mathbb{R}}$ est un point d'accumulation d'une suite $(x_n)_{n \geq 0}$ finie ou infinie de réel, dans $\overline{\mathbb{R}}$, si et seulement si il existe une sous-suite $(x_{n(k)})_{k \geq 0}$ de $(x_n)_{n \geq 0}$ telle que $x_{n(k)}$ converge vers ℓ , quand $k \rightarrow +\infty$.

Exercise 1 : Soient les ensembles $y_n = \inf_{p \geq n} x_p$ et $z_n = \sup_{p \geq n} x_p$ définis pour tout $n \geq 0$. Montrer que :

(1) $\forall n \geq 0, y_n \leq x_n \leq z_n$

(2) Justifier l'existence de la limite de y_n appelée limite inférieure de la suite $(x_n)_{n \geq 0}$, définie par $\liminf x_n$ ou $\underline{\lim} x_n$, et égale à l'expression ci-dessous

$$\underline{\lim} x_n = \liminf x_n = \sup_{n \geq 0} \inf_{p \geq n} x_p$$

(3) Justifier l'existence de la limite de z_n appelée limite supérieure de la suite $(x_n)_{n \geq 0}$ définie par $\limsup x_n$ ou $\overline{\lim} x_n$, et elle est égale à

$$\overline{\lim} x_n = \limsup x_n = \inf_{n \geq 0} \sup_{p \geq n} x_p$$

(4) Etablir que

$$-\liminf x_n = \limsup(-x_n) \quad \text{et} \quad -\limsup x_n = \liminf(-x_n).$$

(5)

Montrer que la limite supérieure est sous-additive et la limite inférieure est sur-additive, i.e. : pour les deux suites $(s_n)_{n \geq 0}$ et $(t_n)_{n \geq 0}$

$$\limsup(s_n + t_n) \leq \limsup s_n + \limsup t_n$$

et

$$\liminf(s_n + t_n) \leq \liminf s_n + \liminf t_n$$

(6) Dédurre de (1) que si

$$\liminf x_n = \limsup x_n,$$

alors $(x_n)_{n \geq 0}$ a une limite et

$$\lim x_n = \liminf x_n = \limsup x_n$$

Exercice 2. Points d'accumulation de la suite $(x_n)_{n \geq 0}$.

(a) Montrer que si $\ell_1 = \liminf x_n$ et $\ell_2 = \limsup x_n$ sont des points d'accumulation $(x_n)_{n \geq 0}$. Montrer un cas et d'en déduire le second en utilisant le point(3) de l'exercice 1.

(b) Montrer que ℓ_1 est le point d'accumulation plus petit $(x_n)_{n \geq 0}$ et ℓ_2

est le plus grand. (De même, montrer un cas et d'en déduire le second en utilisant le point(3) de l'exercice 1).

(c) Dédurre de (a) que si $(x_n)_{n \geq 0}$ a une limite ℓ , alors elle est égal au point d'accumulation unique et donc,

$$\ell = \overline{\lim} x_n = \limsup x_n = \inf_{n \geq 0} \sup_{p \geq n} x_p.$$

(d) Combinez ce résultat au point(6) de l'exercice 1 pour montrer

qu'une suite $(x_n)_{n \geq 0}$ de $\overline{\mathbb{R}}$ a une limite ℓ in $\overline{\mathbb{R}}$ si et seulement si $\liminf x_n = \limsup x_n$ et donc

$$\ell = \lim x_n = \liminf x_n = \limsup x_n$$

Exercice 3. Soit $(x_n)_{n \geq 0}$ une suite croissante de $\overline{\mathbb{R}}$. Etudier sa limite supérieure et la limite inférieure et en déduire que

$$\lim x_n = \sup_{n \geq 0} x_n.$$

En déduire que pour une séquence non croissante $(x_n)_{n \geq 0}$ de $\overline{\mathbb{R}}$,

$$\lim x_n = \inf_{n \geq 0} x_n.$$

Exercice 4. (Les critères de convergence)

Critère 1 . Soit $(x_n)_{n \geq 0}$ une suite réelle de $\overline{\mathbb{R}}$ et un nombre réel $\ell \in \overline{\mathbb{R}}$ tel que: Pour toute sous-suite de $(x_n)_{n \geq 0}$ a également une sous-suite(c'est une sous-suite de $(x_n)_{n \geq 0}$) qui converge vers ℓ . Alors, la limite de $(x_n)_{n \geq 0}$ existe et est égale à ℓ .

Critère 2. Intersections maximales et intersections minimales.

Soit $(x_n)_{n \geq 0}$ une suite dans $\overline{\mathbb{R}}$ et deux nombres réels a et b tel que $a < b$.

Nous définissons

$$\nu_1 = \begin{cases} \inf & \{n \geq 0, x_n < a\} \\ +\infty & \text{if } (\forall n \geq 0, x_n \geq a) \end{cases} .$$

Si ν_1 est fini , soit

$$\nu_2 = \begin{cases} \inf & \{n > \nu_1, x_n > b\} \\ +\infty & \text{if } (n > \nu_1, x_n \leq b) \end{cases} .$$

. Tant que les ν'_j s sont finis, nous pouvons définir pour $\nu_{2k-2} (k \geq 2)$

$$\nu_{2k-1} = \begin{cases} \inf & \{n > \nu_{2k-2}, x_n < a\} \\ +\infty & \text{if } (\forall n > \nu_{2k-2}, x_n \geq a) \end{cases} .$$

et pour ν_{2k-1} fini,

$$\nu_{2k} = \begin{cases} \inf & \{n > \nu_{2k-1}, x_n > b\} \\ +\infty & \text{si } (n > \nu_{2k-1}, x_n \leq b) \end{cases} .$$

Nous nous arrêtons une fois qu'un ν_j est $+\infty$. Si ν_{2j} est fini, alors

$$x_{\nu_{2j}} - x_{\nu_{2j-1}} > b - a.$$

On dit alors : par ce passage de $x_{\nu_{2j-1}}$ à $x_{\nu_{2j}}$, nous avons réalisé un passage (vers le haut) du segment $[a, b]$ appelé *up-crossings*. De même,

si l'un des ν_{2j+1} est fini, alors le segment $[x_{\nu_{2j}}, x_{\nu_{2j+1}}]$ est une intersection de point minimal (downcrossing) pour le segment $[a, b]$. Soit

$$D(a, b) = \text{nombre d'intersection de points maximums } [a, b].$$

- (a) Quelle est la valeur de $D(a, b)$ si ν_{2k} est fini et ν_{2k+1} infini.
- (b) Quelle est la valeur de $D(a, b)$ si ν_{2k+1} est fini et ν_{2k+2} infini.
- (c) Quelle est la valeur de $D(a, b)$ si tous les ν'_j s sont finis.
- (d) Montrer que $(x_n)_{n \geq 0}$ admet une limite si pour tout $a < b$, $D(a, b) < \infty$.
- (e) Montrer que $(x_n)_{n \geq 0}$ admet une limite si pour tout $a < b$, $(a, b) \in \mathbb{Q}^2$, $D(a, b) < \infty$.

Exercice 5. (Critère de Cauchy). Soit $(x_n)_{n \geq 0} \subset \mathbb{R}$ une suite de (nombres réels).

- (a) montrer que si $(x_n)_{n \geq 0}$ est de Cauchy, alors elle admet un point accumulation unique $\ell \in \mathbb{R}$ qui est sa limite.
- (b) Montrer que si une suite $(x_n)_{n \geq 0} \subset \mathbb{R}$ converge vers $\ell \in \mathbb{R}$, alors, elle est de Cauchy.
- (c) Dédire le critère de Cauchy pour les suites de nombres réels.

SOLUTIONS

Exercice 1.

Question (1) : Il est évident que :

$$\inf_{p \geq n} x_p \leq x_n \leq \sup_{p \geq n} x_p,$$

étant donné que x_n est un élément de $\{x_n, x_{n+1}, \dots\}$ sur laquelle nous prenons le supremum ou infimum.

Question (2) : Soit $y_n = \inf_{p \geq 0} x_p = \inf_{p \geq n} A_n$, o $A_n = \{x_n, x_{n+1}, \dots\}$ est une suite non croissante de l'ensemble : $\forall n \geq 0$,

$$A_{n+1} \subset A_n.$$

Ainsi l'infimum dans A_n est croissant. Si y_n est croissante dans $\overline{\mathbb{R}}$, sa limite est au dessus des bornes finie ou infinie. Ainsi

$$y_n \nearrow \underline{\lim} x_n,$$

est un nombre fini ou infini.

Question (3) : Nous démontrons aussi que $z_n = \sup A_n$ décroît et $z_n \downarrow \overline{\lim} x_n$.

Question (4) : Nous rappelons que

$$-\sup \{x, x \in A\} = \inf \{-x, x \in A\}.$$

lequel , nous permet d'écrire

$$-\sup A = \inf -A.$$

Ainsi,

$$-z_n = -\sup A_n = \inf -A_n = \inf \{-x_p, p \geq n\} ..$$

Le terme de la droite tend vers $-\overline{\lim} x_n$ et celui de la gauche vers $\underline{\lim} -x_n$ et donc

$$-\overline{\lim} x_n = \underline{\lim} (-x_n).$$

De même, nous montrons que:

$$-\underline{\lim} (x_n) = \overline{\lim} (-x_n).$$

Question (5). Ces propriétés viennent des formules suivante ,o $A \subseteq \mathbb{R}, B \subseteq \mathbb{R}$:

$$\sup \{x + y, A \subseteq \mathbb{R}, B \subseteq \mathbb{R}\} \leq \sup A + \sup B.$$

De ce fait :

$$\forall x \in \mathbb{R}, x \leq \sup A$$

et

$$\forall y \in \mathbb{R}, y \leq \sup B.$$

Ainsi,

$$x + y \leq \sup A + \sup B,$$

o

$$\sup_{x \in A, y \in B} x + y \leq \sup A + \sup B.$$

De même,

$$\inf(A + B) \geq \inf A + \inf B.$$

De ce fait:

$$\forall (x, y) \in A \times B, x \geq \inf A \text{ and } y \geq \inf B.$$

Ainsi,

$$x + y \geq \inf A + \inf B.$$

Ains,

$$\inf_{x \in A, y \in B} (x + y) \geq \inf A + \inf B$$

Application.

$$\sup_{p \geq n} (x_p + y_p) \leq \sup_{p \geq n} x_p + \sup_{p \geq n} y_p.$$

Toutes ces suites sont non croissantes. Prenant l'infimum, nous obtenons la limite superieur:

$$\overline{\lim} (x_n + y_n) \leq \overline{\lim} x_n + \overline{\lim} y_n.$$

Question (6) : Mettons

$$\underline{\lim} x_n = \overline{\lim} x_n,$$

Comme :

$$\forall x \geq 1, y_n \leq x_n \leq z_n,$$

$$y_n \rightarrow \underline{\lim} x_n$$

et

$$z_n \rightarrow \overline{\lim} x_n,$$

Nous appliquons le théorème de Sandwich pour conclure que la limite de x_n existe et :

$$\lim x_n = \underline{\lim} x_n = \overline{\lim} x_n.$$

Exercice 2.

Question (a).

La question (4) de l'exercice 1 étant prouvée, il suffit maintenant de montrer cette propriété pour l'une des limites. Considérons la limite supérieure et les trois cas d'étude suivants:

le cas de la limite supérieure finie :

$$\underline{\lim} x_n = \ell \text{ finie.}$$

Par définition,

$$z_n = \sup_{p \geq n} x_p \downarrow \ell.$$

Donc:

$$\forall \varepsilon > 0, \exists (N(\varepsilon) \geq 1), \forall p \geq N(\varepsilon), \ell - \varepsilon < x_p \leq \ell + \varepsilon.$$

Prenons le moins que celui là :

$$\forall \varepsilon > 0, \exists n_\varepsilon \geq 1 : \ell - \varepsilon < x_{n_\varepsilon} \leq \ell + \varepsilon.$$

Nous pouvons construire une sous-suite convergeant vers ℓ .

Soit $\varepsilon = 1$:

$$\exists N_1 : \ell - 1 < x_{N_1} = \sup_{p \geq N_1} x_p \leq \ell + 1.$$

Mais si

$$(1.1) \quad z_{N_1} = \sup_{p \geq N_1} x_p > \ell - 1,$$

il existe sûrement un $n_1 \geq N_1$ tel que

$$x_{n_1} > \ell - 1.$$

sinon nous nous pourrions avoir

$$(\forall p \geq N_1, x_p \leq \ell - 1) \implies \sup \{x_p, p \geq N_1\} = z_{N_1} \leq \ell - 1,$$

lequel est contraire à (1.1). Donc, il existe $n_1 \geq N_1$ tel que

$$\ell - 1 < x_{n_1} \leq \sup_{p \geq N_1} x_p \leq \ell - 1.$$

i.e.

$$\ell - 1 < x_{n_1} \leq \ell + 1.$$

Nous progressons avec l'étape $\varepsilon = \frac{1}{2}$ et nous considérons la suite $(z_n)_{n \geq n_1}$ dont la limite reste égale à ℓ . Donc, il existe $N_2 > n_1$:

$$\ell - \frac{1}{2} < z_{N_2} \leq \ell - \frac{1}{2}.$$

Nous déduisons comme précédemment que $n_2 \geq N_2$ tel que

$$\ell - \frac{1}{2} < x_{n_2} \leq \ell + \frac{1}{2}$$

Avec $n_2 \geq N_1 > n_1$.

Ensuite, nous mettons $\varepsilon = 1/3$, il existera $N_3 > n_2$ tel que

$$\ell - \frac{1}{3} < z_{N_3} \leq \ell - \frac{1}{3}$$

et nous pourrons voir un $n_3 \geq N_3$ tel que

$$\ell - \frac{1}{3} < x_{n_3} \leq \ell - \frac{1}{3}.$$

Pas à pas, nous déduisons l'existence de $x_{n_1}, x_{n_2}, x_{n_3}, \dots, x_{n_k}, \dots$ avec $n_1 < n_2 < n_3 < \dots < n_k < n_{k+1} < \dots$ tel que

$$\forall k \geq 1, \ell - \frac{1}{k} < x_{n_k} \leq \ell - \frac{1}{k},$$

i.e.

$$|\ell - x_{n_k}| \leq \frac{1}{k}.$$

Ce qui implique que:

$$x_{n_k} \rightarrow \ell$$

Conclusion : $(x_{n_k})_{k \geq 1}$ est bien une sous-suite tel que $n_k < n_{k+1}$ pour tout $k \geq 1$ et il converge vers ℓ , qui est alors un point d'accumulation.

Le cas de la limite supérieure égale à $+\infty$:

$$\overline{\lim} x_n = +\infty.$$

Lorsque $z_n \uparrow +\infty$, nous avons : $\forall k \geq 1, \exists N_k \geq 1$,

$$z_{N_k} \geq k + 1.$$

Pour $k = 1$, soit $z_{N_1} = \inf_{p \geq N_1} x_p \geq 1 + 1 = 2$. Donc il existe

$$n_1 \geq N_1$$

tel que :

$$x_{n_1} \geq 1.$$

Pour $k = 2$: considerons les suites $(z_n)_{n \geq n_1+1}$. De la même manière, nous trouvons

$$n_2 \geq n_1 + 1$$

et

$$x_{n_2} \geq 2.$$

Etape par étape, nous trouvons pour tout $k \geq 3$, et $n_k \geq n_{k-1} + 1$ tel que

$$x_{n_k} \geq k.$$

Ce qui conduit à $x_{n_k} \rightarrow +\infty$ comme $k \rightarrow +\infty$.

Cas de la limit superieur égale à $-\infty$:

$$\overline{\lim} x_n = -\infty.$$

Cela implique : $\forall k \geq 1, \exists N_k \geq 1$, tel que

$$z_{n_k} \leq -k.$$

Pour $k = 1, \exists n_1$ tel que

$$z_{n_1} \leq -1.$$

Mais

$$x_{n_1} \leq z_{n_1} \leq -1$$

Soit $k = 2$. Considerons $(z_n)_{n \geq n_1+1} \downarrow -\infty$. Il existera $n_2 \geq n_1 + 1$:

$$x_{n_2} \leq z_{n_2} \leq -2$$

Etape par étape, nous trouvons $n_{k+1} < n_k$ de telle sorte que $x_{n_k} < -k$ pour tout k plus grand que 1. Donc,

$$x_{n_k} \rightarrow +\infty$$

Question (b).

Soit ℓ un point d'accumulation de $(x_n)_{n \geq 1}$, la limite d'un de ses suites $(x_{n_k})_{k \geq 1}$. Nous avons

$$y_{n_k} = \inf_{p \geq n_k} x_p \leq x_{n_k} \leq \sup_{p \geq n_k} x_p = z_{n_k}$$

Le terme de gauche est une sous suite de (y_n) tendant vers la limite inférieure et le côté droit est une sous suite de (z_n) tendant vers la limitesuperieur. Donc, nous obtenons :

$$\underline{\lim} x_n \leq \ell \leq \overline{\lim} x_n,$$

ce qui montre que $\liminf x_n$ est le plus petit point d'accumulation et $\overline{\lim} x_n$ est le plus grand.

Question (c). Si la suite $(x_n)_{n \geq 1}$ admet une limite ℓ , alors cette limite est celle de toutes ses sous-suites, donc les sous-suite tend vers les limites supérieur ou inférieur. Ce qui répond à la question (b).

Question (d). Nous répondons à cette question en combinant le point (d) de cette exercice et le point (6) de l'exercice 1.

Exercice 3. Soit $(x_n)_{n \geq 0}$ une suite non-décroissante, nous avons:

$$z_n = \sup_{p \geq n} x_p = \sup_{p \geq 0} x_p, \forall n \geq 0.$$

Pourquoi? Parce que par la croissance nous avons,

$$\{x_p, p \geq 0\} = \{x_p, 0 \leq p \leq n-1\} \cup \{x_p, p \geq n\}$$

Etant donné que tous les éléments de $\{x_p, 0 \leq p \leq n-1\}$ sont plus petits que ceux de $\{x_p, p \geq n\}$, le supremum est atteint en $\{x_p, p \geq n\}$ et donc

$$\ell = \sup_{p \geq 0} x_p = \sup_{p \geq n} x_p = z_n$$

Ainsi

$$z_n = \ell \rightarrow \ell.$$

Nous avons aussi $y_n = \inf \{x_p, 0 \leq p \leq n\} = x_n$ qui est une suite non-décroissante et ainsi converge vers $\ell = \sup_{p \geq 0} x_p$.

Exercice 4.

Soit $\ell \in \overline{\mathbb{R}}$ ayant la propriété indiquée. Soit ℓ' un point d'accumulation donné.

$$(x_{n_k})_{k \geq 1} \subseteq (x_n)_{n \geq 0} \text{ tel que } x_{n_k} \rightarrow \ell'.$$

Par hypothèse cette suite (x_{n_k}) a à son tour une sous-suite $(x_{n_{(k(p))}})_{p \geq 1}$ tel que $x_{n_{(k(p))}} \rightarrow \ell$ as $p \rightarrow +\infty$.

Mais comme une sous-suite de $(x_{n_{(k)}})$,

$$x_{n_{(k(\ell))}} \rightarrow \ell'.$$

Ainsi

$$\ell = \ell'.$$

En appliquant cela à la limite supérieure et à la limite inférieure, nous avons:

$$\overline{\lim} x_n = \underline{\lim} x_n = \ell.$$

Et donc $\lim x_n$ existe et est égal à ℓ .

Exercice 5.

Question (a). Si ν_{2k} fini et ν_{2k+1} infini, il a alors exactement k un point d'intersection supérieur: $[x_{\nu_{2j-1}}, x_{\nu_{2j}}]$, $j = 1, \dots, k : D(a, b) = k$.

Question (b). Si ν_{2k+1} fini et ν_{2k+2} infini, il a alors exactement k un point d'intersection supérieur: $[x_{\nu_{2j-1}}, x_{\nu_{2j}}]$, $j = 1, \dots, k : D(a, b) = k$.

Question (c). Si tous les ν'_j s sont finis, alors, il existent un nombre infini de points d'intersection supérieur: $[x_{\nu_{2j-1}}, x_{\nu_{2j}}]$, $j \geq 1 : D(a, b) = +\infty$.

Question (d). Supposons qu'il existe $a < b$ des nombres rationnels tels que $D(a, b) = +\infty$. Alors tous les ν'_j s sont finis. La sous-suite $x_{\nu_{2j-1}}$ est strictement inférieure a . Donc, sa limite inférieure est inférieure à a . Cette limite inférieure est un point d'accumulation de la suite $(x_n)_{n \geq 1}$, il en est de plus $\underline{\lim} x_n$, qui est inférieur à a .

De même, la sous-suite $x_{\nu_{2j}}$ est strictement inférieure b . Donc, la limite supérieure est au-dessus de a . Cette limite supérieure est un point d'accumulation de la suite $(x_n)_{n \geq 1}$, donc elle est inférieure à $\overline{\lim} x_n$, qui est directement au-dessus b . Qui conduit à:

$$\underline{\lim} x_n \leq a < b \leq \overline{\lim} x_n.$$

Cela implique que la limite de (x_n) n'existe pas. En revanche, nous venons de prouver que la limite de (x_n) existe, Entre-temps pour tous les nombres réels a et b tel que $a < b$, $D(a, b)$ est fini.

Maintenant, supposons que la limite de (x_n) n'existe pas. Alors,

$$\underline{\lim} x_n < \overline{\lim} x_n.$$

Nous pouvons alors trouver deux rationnels a et b tel que $a < b$ et un nombre ϵ tel que $0 < \epsilon$, tous les

$$\underline{\lim} x_n < a - \epsilon < a < b < b + \epsilon < \overline{\lim} x_n.$$

Si $\underline{\lim} x_n < a - \epsilon$, nous pouvons revenir de la question (a) de exercice 2 et construire une suite (x_n) ce qui tend vers $\underline{\lim} x_n$ tout en restant en dessous de $a - \epsilon$. De même, si $b + \epsilon < \overline{\lim} x_n$, nous pouvons créer une suite de (x_n) ce qui tend vers $\overline{\lim} x_n$ tout en restant au-dessus $b + \epsilon$. Il est évident avec ces deux suites que nous pourrions définir avec ces deux suites tous ν_j fini et donc $D(a, b) = +\infty$.

Nous venons de montrer par l'absurde que si tous les $D(a, b)$ sont finis pour tous les rationnels a et b tel que $a < b$, alors, la limite de (x_n) existe.

Exercice 5. critère de Cauchy dans \mathbb{R} .

Supposons que la suite est Cauchy, *i.e.*,

$$\lim_{(p,q) \rightarrow (+\infty, +\infty)} (x_p - x_q) = 0.$$

Ensuite, soient $x_{n_{k,1}}$ et $x_{n_{k,2}}$ deux sous-suites convergeant respectivement vers $\ell_1 = \underline{\lim} x_n$ et $\ell_2 = \overline{\lim} x_n$. Alors

$$\lim_{(p,q) \rightarrow (+\infty, +\infty)} (x_{n_{p,1}} - x_{n_{q,2}}) = 0.$$

En première location $p \rightarrow +\infty$, nous avons

$$\lim_{q \rightarrow +\infty} \ell_1 - x_{n_{q,2}} = 0.$$

Ce qui montre que la limite ℓ_1 est finie, sinon $\ell_1 - x_{n_{q,2}}$ demeurerait infinie et ne tendent pas vers 0. En échangeant les rôles de p et q , nous avons aussi que ℓ_2 est fini.

Enfin, en laissant $q \rightarrow +\infty$, dans la dernière équation, nous obtenons

$$\ell_1 = \underline{\lim} x_n = \overline{\lim} x_n = \ell_2.$$

ce qui prouve l'existence d'une limite finie de la suite (x_n) .

Supposons maintenant que la limite finie ℓ de (x_n) existe. Alors

$$\lim_{(p,q) \rightarrow (+\infty, +\infty)} (x_p - x_q) = \ell - \ell = 0.$$

0 Ce qui montre que la suite est de Cauchy.

2. Miscelleanuous facts

LEMME 11. . For any $a \in \mathbb{R}$,

$$|e^{ia} - 1| = \sqrt{2(1 - \cos a)} \leq 2 |\sin(a/2)| \leq 2 |a/2|^\delta.$$

Proof. This is easy for $|a/2| > 1$. Indeed for $\delta > 0$, $|a/2|^\delta > 0$ and

$$2 |\sin(a/2)| \leq 2 \leq 2 |a/2|^\delta$$

Now for $|a/2| > 1$, we have the expansion

$$\begin{aligned} 2(1 - \cos a) &= a^2 - \sum_{k=2}^{\infty} (-1)^2 \frac{a^{2k}}{(2k)!} = x^2 - 2 \sum_{k \geq 2, k \text{ even}}^{\infty} \frac{a^{2k}}{(2k)!} - \frac{a^{2(k+1)}}{(2(k+1))!} \\ &= a^2 - 2x^{2(k+1)} \sum_{k \geq 2, k \text{ even}}^{\infty} \frac{1}{(2k)!} \left\{ \frac{1}{a^2} - \frac{1}{(2k+1)((2k+2)\dots(2k+k))} \right\}. \end{aligned}$$

For each $k \geq 2$, for $|a/2| < 1$,

$$\left\{ \frac{1}{a^2} - \frac{1}{(2k+1)((2k+2)\dots(2k+k))} \right\} \geq \left\{ \frac{1}{4} - \frac{1}{(2k+1)((2k+2)\dots(2k+k))} \right\} \geq 0.$$

Hence

$$2(1 - \cos a) \leq a^2.$$

But for $|a/2|$, the function $\delta \mapsto |a/2|^\delta$ is non-increasing $\delta, 0 \leq \delta \leq 1$.

Then

$$\sqrt{2(1 - \cos a)} \leq |a| = 2 |a/2|^1 \leq 2 |a/2|^\delta.$$

Bibliography

- [1] Bauer, H.(1981). *Probability Theory and Elements of Measure Theory*. Holt, Rinehart, and Winston, New-York.
- [2] Billingsley, P.(1968). *Convergence of Probability measures*. John Wiley, New-York.
- [3] Dudley, R. M.(1989). *Real Analysis and probability*. Wadsworth, Pacific Grove.
- [4] Gutt, A.(2005). *Probability : A Graduate Course*. Springer-Verlag.
- [5] Lo, G.S.(2016). *A Course on Elementary Probability Theory*. SPAS Editions. Saint-Louis, Calgary, Abuja. Doi : 10.16929/sbs/2016.0003
- [6] Lo, G.S.(2016). *Cours Elementaire de Théorie de Probabilités*. SPAS Editions. Saint-Louis, Calgary, Abuja. Doi : 10.16929/sbs/2016.0004
- [7] Lo, G.S.(2016). *Introduction to stochastic processes*. Spas Textbooks Series.
- [8] Lo, G.S.(2016). *Mathematical Foundation to Probability Theory*. Spas Textbooks Series.
- [9] Loève, Michel.(1997). *Probability Theory I*. Springer-Verlag, 4th Edition.
- [10] A. W. van der Vaart and J. A. Wellner(1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, New-York.
- [11] van der Vaart, A.W. *Asymptotics Statistics*. (2000). *Cambridge*